

Access DB# 85268**SEARCH REQUEST FORM**

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

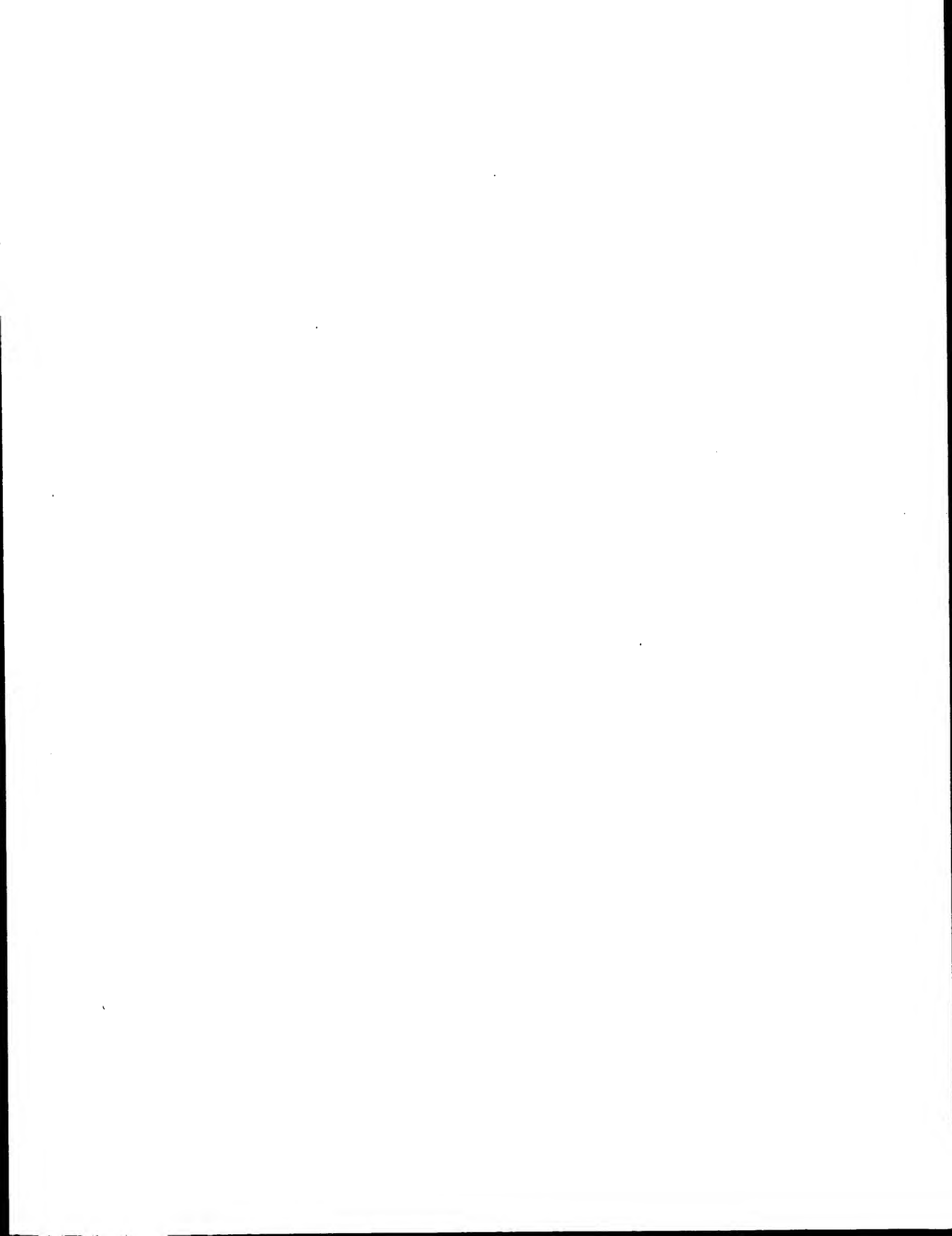
**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 - 703-308-4498
jan.delaval@uspto.gov

STAFF USE ONLY**Type of Search****Vendors and cost where applicable**

Searcher: Ch
Searcher Phone #: 4458
Searcher Location: _____
Date Searcher Picked Up: 1/22/03
Date Completed: 2/2/03
Searcher Prep & Review Time: _____
Clerical Prep Time: 30
Online Time: 75

NA Sequence (#) <input checked="" type="checkbox"/>	STN _____
AA Sequence (#) <input checked="" type="checkbox"/>	Dialog _____
Structure (#) _____	Questel/Orbit _____
Bibliographic _____	Dr.Link _____
Litigation _____	Lexis/Nexis _____
Fulltext _____	Sequence Systems <input checked="" type="checkbox"/>
Patent Family _____	WWW/Internet _____
Other _____	Other (specify) _____



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:23 ; Search time 26.5 Seconds
(without alignments)
1749.457 Million cell updates/sec

Title: US-09-931-704-2
Perfect score: 1226
Sequence: 1 MDLRAGDSWGLACLCTVLW.....KKKMQPPAAAVTLHLGAHGF 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp Vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1226	100.0	225	4	Q9UBD9
2	1193	97.3	225	11	Q9QZM3
3	150.5	12.3	215	13	Q9PUJ2
4	150.5	12.3	215	13	Q9PUJ1
5	150.5	12.3	215	13	Q9PUJ0
6	148.5	12.1	215	13	Q9PUJ9
7	97.5	8.0	318	4	Q96LS2
8	97.5	8.0	530	3	Q8X0E9
9	96.5	7.9	455	11	Q9CWV7
10	92	7.5	332	10	Q9MAV1
11	92	7.5	423	11	Q9JHE4
12	90	7.3	423	11	Q9D8V6
13	89.5	7.3	771	2	Q9S3Q9
14	89	7.3	733	16	Q9I664
15	87	7.1	955	11	O88287
16	87	7.1	1561	11	O88286

ALIGNMENTS

RESULT 1

ID	Q9UBD9	PRELIMINARY;	PRT;	225 AA.
AC	Q9UBD9;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like cytokine) (Similar to cardiotrophin-like cytokine, neurotrophin-1/B-cell stimulating factor-3).			
DE	CtC.			
GN	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI_TaxID=9606;				
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99432254; PubMed=10500198;			
RA	Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S., Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F., Simonet W.S., Boone T., Chang M.-S.;			
RA	"Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the IL-6 family."			
RT	Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=99382254; PubMed=10448081;			
RX	Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J., Ruben S., Alderson R.F.;			
RA	"Computational EST database analysis identifies a novel member of the neurotrophic cytokine family."			
RT	Biochem. Biophys. Res. Commun. 262:132-138(1999).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;			
RA	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=KIDNEY;			

O54153 streptomyc
O921r2 mus musculu
O8363 mus musculu
O8841 mus musculu
Q9cp21 mus musculu
Q9ya71 aeropyrum p
Q8ybb1 brucella me
Q9nxx5 homo sapien
Q9rvf3 deinococcus
O8432 treponema p
Q9p210 homo sapien
Q95159 homo sapien
Q9rlp6 mycobacteri
Q91584 streptococc
Q9uid0 homo sapien
Q9nbs5 pseudomonas
Q9287 pseudomonas
Q981p5 rhizobium l
Q96pc8 homo sapien
Q96pc7 homo sapien
O6507 arabidopsis
Q9fin7 arabidopsis
Q9a100 frankia sp.
Q9e122 human immun
Q8ug24 rhodobacteri
Q9x582 rhodothermu
Q9h7r6 homo sapien
Q8x0g6 anabaena sp

85.5 7.1 640 16 O54153
85.5 7.0 1154 11 Q921r2
85 6.9 476 11 Q8363
85 6.9 727 11 O8841
21 84.5 6.9 294 11 Q9CP21
22 84.5 6.9 389 17 Q9YA71
23 84 6.9 200 16 Q8YBB1
24 84 6.9 294 4 Q9NXX5
25 84 6.9 411 16 Q9RVF3
26 84 6.9 559 16 O8432
27 84 6.9 876 4 Q9P210
28 83.5 6.8 310 4 Q95159
29 83.5 6.8 5990 2 Q9RLP6
30 83 6.8 222 2 Q91584
31 82.5 6.7 310 4 Q9UID0
32 82.5 6.7 542 16 Q9HTB5
33 82 6.7 283 2 Q9ZB87
34 82 6.7 611 16 Q98LP5
35 82 6.7 721 4 Q96PC8
36 82 6.7 756 4 Q96PC7
37 82 6.7 1179 10 O65507
38 82 6.7 2376 10 Q9FIN7
39 81.5 6.6 396 2 Q9AF00
40 81 6.6 328 15 Q9EI25
41 81 6.6 328 15 Q9EI22
42 81 6.6 479 16 O8UG24
43 81 6.6 924 2 Q9X582
44 80.5 6.6 305 4 Q9H7R6
45 80.5 6.6 353 16 Q8X0G6

```

RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DDSI databases.
DR EMBL; AF176912; AAF00992.1; -
DR EMBL; AF172854; AAD54284.1; -
DR EMBL; AF176911; AAF00991.1; -
DR EMBL; AY049779; AAL15436.1; -
DR EMBL; BC012939; AAH12939.1; -
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Query Match 100.0%; Score 1226; DB 4; Length 225;
Best Local Similarity 100.0%; Pred. No. 3.6e-106;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMGLACTVWLHPVAVPALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVWLHPVAVPALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGT 60

Qy 61 YLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL 120
Db 61 YLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTTPGPAHSDFLQ 180
Db 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPAAAVTLHLGAHGF 225

RESULT 2
Q9Q2M3 PRELIMINARY; PRT; 225 AA.
ID Q9Q2M3 AC Q9Q2M3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Neurotrophin-1/B-cell stimulating factor-3.
GN BSF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eumariota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
RT IL-6 family";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463 (1999).
DR EMBL; AF176913; AAF00993.1; -
DR MGD; MGI:1930088; Bsf3.
SQ SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;

Query Match 97.3%; Score 1193; DB 11; Length 225;
Best Local Similarity 96.9%; Pred. No. 4.2e-103;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMGLACTVWLHPVAVPALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVWLHPVAVPALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGT 60

Qy 61 YLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL 120
Db 61 YLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTTPGPAHSDFLQ 180
Db 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPAAAVTLHLGAHGF 225

RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DDSI databases.
DR EMBL; AF176912; AAF00992.1; -
DR EMBL; AF172854; AAD54284.1; -
DR EMBL; AF176911; AAF00991.1; -
DR EMBL; AY049779; AAL15436.1; -
DR EMBL; BC012939; AAH12939.1; -
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Query Match 100.0%; Score 1226; DB 4; Length 225;
Best Local Similarity 100.0%; Pred. No. 3.6e-106;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMGLACTVWLHPVAVPALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVWLHPVAVPALNRTGDPGPGPSIQKTYDLYLRYLHQLRSLAGT 60

Qy 61 YLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL 120
Db 61 YLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTTPGPAHSDFLQ 180
Db 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPAAAVTLHLGAHGF 225

RESULT 3
Q9PUJ2 PRELIMINARY; PRT; 215 AA.
ID Q9PUJ2 AC Q9PUJ2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Receptivity factor isoform 1 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
RT salamander.";
RL Science 285:1907-1909 (1999).
DR EMBL; AF181480; AAF01025.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 215 AA; 24138 MW; B1906BB666335738 CRC64;

Query Match 12.3%; Score 150.5; DB 13; Length 215;
Best Local Similarity 26.2%; Pred. No. 3.8e-06;
Matches 42; Conservative 31; Mismatches 80; Indels 7; Gaps 3;

Qy 56 SLACTYLYNLGPPNEPDPNPRGLAETLPRATVDLEWVRSNDKRLRTQNYEAYSHLLC 115
Db 55 SLLPTLYLSFQCAPLSDPDYQPLPHKLVANPLRAMDYDFMKQTDRETLNNNLYFYSAIVE 114

Qy 116 YLR-GLNRQ----AATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTPTTWT 170
Db 115 FLKEAMTQEDLNPAELSLKAKFEAMANSNTLISKISDINTQMGMSVTTILP--KPLVV 172

Qy 171 PGPAHSDFLQMDDFWLLKELQTLWLRSAKDFNRLKKMKQ 210
Db 173 PFEGSAVPRKKLGGVGVCKEYKERVLLTKRDFEFLAKKYQ 212

RESULT 4
Q9PUJ1 PRELIMINARY; PRT; 215 AA.
ID Q9PUJ1 AC Q9PUJ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Receptivity factor isoform 2 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
RT salamander.";
RL Science 285:1907-1909 (1999).
DR EMBL; AF181481; AAF01026.1; -
KW Signal.
FT SIGNAL.
SQ SEQUENCE 215 AA; 24080 MW; B341B8B7B4E28438 CRC64;

```


Db 235 LFGSRSTPSTPVES 249

RESULT 8

Q8X0E9 PRELIMINARY; PRT; 530 AA.

AC Q8X0E9; (TREMELrel. 20, Created)

DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Hypothetical 59.4 kDa protein.

GN B14A6.080.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

EN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.,

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL; AL670007; CAD21310.1; -.

DR InterPro; IPR001810; F-box.

DR PROSITE; PS0181; FBOX; 1.

KW Hypothetical protein.

SQ SEQUENCE 530 AA; 59352 MW; 70382EEL5F71BB9D CRC64;

Query Match 8.0%; Score 97.5; DB 3; Length 530;

Best Local Similarity 24.8%; Pred. No. 1;

Matches 41; Conservative 22; Mismatches 65; Indels 37; Gaps 6;

QY 64 YLGPPEPPENPRLGAETLPA-TVDLEWRSINDKRLRLTQNYEAYSHLLCYLRLNLR 122

Db 236 FLVPPLEPDPFLVLAGLKKLHLVDADLATHRDMEIQYLLFKKFLALTPNLTLWL 291

QY 123 QAATAELRRSLAHFCTSLQGLGSIAGMAALGYPLQPLPGT-----EPTWTPGP 173

Db 292 ----VNPERISISHY-----GGKEALLKWLASSLREGTWSPPGLINADPSRLPPP 336

QY 174 AHSDFLQKMDFFWLKELQTLWPSAKDFNRLKKMQPPAAAVTL 218

Db 337 VEFHLEQLD-----IGQLDV-----SANTLYLRFNKSSTLKAISL 373

RESULT 9

Q9CWV7 PRELIMINARY; PRT; 455 AA.

AC Q9CWV7; (TREMELrel. 17, Created)

DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)

DE 2410003H12Rik protein.

GN 2410003H12Rik.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC STEM CELLS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staabli F., Suzuki R., Tomita R., Wagner L., Washio T.,

Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momabaerts P.,

Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,

Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690 (2001).

DR EMBL; AK010358; BAB26878.1; -.

DR MGI; MGI:1919221; 2410003H12Rik.

SQ SEQUENCE 455 AA; 51481 MW; 201886B814EB9CFF CRC64;

Query Match 7.9%; Score 96.5; DB 11; Length 455;

Best Local Similarity 22.4%; Pred. No. 1.1;

Matches 59; Conservative 24; Mismatches 89; Indels 91; Gaps 12;

QY 7 DSWGMLACLCTVLW-HLPAVPAALNRTGDPG-----PGPSIOKTYDLTRYLEHQLRS 56

Db 29 DAFNYQSCFMQDLFAHFPEVLFIHRTYNPRGKLVYTLVDGPRVQVEGLARAV----- 82

QY 57 LAGTYLNYLGPFFNE-----PDFNPPRLGAETLPRATVDLEWRSINDKRLRLT 104

Db 83 -----YFAIPTNEDARGLAQMFQVKKFNP-----AWERNVTIL----- 116

QY 105 QNYEAYSHLLCYLRLNRAQATAELRRSLAHFCTSLQGLGSIAGMAALGYPLQPLPQ 161

Db 117 ----VDPHFL-LLPTLTWEFPTAEVLLSAFHICKFLQCKFYQL-----PLEQFVQR 162

QY 162 ----LPQTEPTWTPGAH-----SDFLQKMDFFWLKELQTLW-----WRSKADF 202

Db 163 LLLSSLQSTMCSATAGNLKLYTLNLCIPSSRLPELHSHWLND-RWLAWHRSRAQS 221

QY 203 NRLKKMQPPAAAVTLHLGAHGF 225

Db 222 SRYFOSLEIMAHILSQFFGTTPF 244

RESULT 10

Q9MAUI PRELIMINARY; PRT; 332 AA.

AC Q9MAUI; (TREMELrel. 15, Created)

DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)

DE F13M7.8 protein.

GN F13M7.8.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu, S.,

Li J., Kremenetskaia I., Luros J., Araujo R., Au M., Bredel V.,

Buehler E., Conway A., Dewar K., Feng J., Kim C., Kim D., Li Y.,

Palm C., Shinn P., Sun H., Davis R., Ecker J., Federspiel N.,

Theologis A.,

RT "The sequence of BAC F13M7 from Arabidopsis thaliana chromosome 1.1";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Theologis A.; (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC004809; AAF40444.1; -.

DR InterPro; IPR002965; P-rich extensin.

DR PRINTS; PR01217; PRICHEXTENSIN.

```

SQ SEQUENCE 332 AA; 36793 MW; 18E8687141A070F4 CRC64;
Query Match
Best Local Similarity 26.4%; Score 92; DB 10; Length 332;
Matches 58; Conservative 24; Mismatches 72; Indels 66; Gaps 13;

QY 23 PAVPALNRTGDPGPGSPSTQKTYDLYLEHQLRSIAGTYLNY-----64
Df 113 PSVTAGNLGSGYP-PPSPF--TYDFGPEQRMESILLOQFIRERNQIRPLRGLGLSPVG 169
QY 65 LGPPFNEPDPNPPRLGAEATLPRATVDLEWVRSNDKRLTQNYEAXSHLLCYLRLGNRQA 124
Df 170 LGPIRASFPQLQVRAPP--PTSILD--TSNRKAR-----SKOGALAVRG--RKV 215
QY 125 ATAELRSL-----AHFCTSLOGLGSIAGVMAALGYPLPQLP---GTEPTWT 170
Df 216 RITEGSSSLYSLGRSLKNGAHV-----GIQPRSGINK-----PLPKPLPVDLTITTSVP 266
QY 171 PGPAHSDFLOKMDDFWLLKELQTLWRSKADPNRLKKQW 210
Df 267 DDFDESADKDEEAVKQL-----SEKDL--LKRHIE 298

RESULT 11
Q9JHE4 PRELIMINARY; PRT; 423 AA.
AC Q9JHE4;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Cerebroside sulfotransferase.
GN GCST OR CST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boiffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007645; BAB25160.1; -.
DR MGD; MGI:1858277; Gcst.
SQ SEQUENCE 423 AA; 48824 MW; 045303CFCE150A3 CRC64;

Query Match
Best Local Similarity 22.7%; Score 90; DB 11; Length 423;
Matches 55; Conservative 22; Mismatches 61; Indels 104; Gaps 12;

QY 44 YDLTRYLEHQLRSIAGTYLNY---LGPP-----FNEP-----72
Df 204 YDPSSYNAHYLRNLFFDLGYDSSLDPASPRVQEHILEVERRHFLVLLQYFDESILVLLQ 263
QY 73 -----DENPPRLGAEATLPRAT-----VDLEWVRSNDKRLTQNYEA 109
Df 264 ELLCWDLEDVLYFKLNARSDSPVRLSGELRYRATAMNLLDVLRYHFN--ASFWRKVEA 321
QY 110 YSHLLCYLRLGNRQA-ATAELRRS---LAHFCTSLOGLGSI-----GVMAA 153
Df 322 F-----GRERMAVEALRQANEHMEHICIDGQAVGAEAIRDSAMQPWQPLGIKSI 373
QY 154 LGYPLPQLPGTEPTWTGPAHSDFLOKMDDFWLLKELQ-----TWLWRSKAD 202
Df 374 LGYNLKKSI-----GPOHEQLCRMLTPETQVLSDLGANLWVTKLWFLRDF 420
QY 203 NR 204
Df 421 LR 422

RESULT 13
Q9S3Q9 PRELIMINARY; PRT; 771 AA.
AC Q9S3Q9
ID Q9S3Q9
DT 01-MAY-2000 (TREMELrel. 13, Created)

```

```

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Immunoreactive 87kd antigen PG92.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "Porphyromonas gingivalis polypeptides and nucleic acids.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF175724; AAD51077.1; -.
DR TIGRFAMs; TIGR01180; aman2; put. 1.
SQ SEQUENCE 771 AA; 86667 MW; 75016B6F6848C9B9 CRC64;

Query Match 7.3%; Score 89.5; DB 2; Length 771;
Best Local Similarity 25.6%; Pred. No. 9.4;
Matches 50; Conservative 16; Mismatches 48; Indels 81; Gaps 13;

QY 9 WGNLACLCITVL-----NHLPAVPAALNRTGDPGPGPSIQKTYDLYRLEHQLRSLA----- 58
DB 549 YGM---LCPLLPGSFLTPDPKQGENFENPFGEHGSAYNYAFFVPHDIQGLARLMGGA 605

QY 59 -----GTYNLVLPFPNPFDPNPRGAEATLPRATVDLEWRSNDKRLQ 105
DB 606 KVFSERLQKVFDEGYV-----DPTNEPDIAYPVL-PSYFFPK-----EAWR----- 644

QY 106 NYEAYSHLLCYLRGLNRQAATLRLSL--AHCTSLQGLLGS-IAGVMAA-----LG- 155
DB 645 -----TQKLTRELIDKHFCNAPNGLPGNDAGTMSAWLVYSMLGF 684

QY 156 YPLPQPLPGTEPTWT 170
DB 685 YP---DCPGS-PIYT 695

RESULT 14
QY1664 PRELIMINARY; PRT; 733 AA.
AC QY1664;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Hypothetical protein PA0454.
GN PA0454.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lam R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004483; AAG03843.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 733 AA; 81538 MW; 580F9DBECB3909DB CRC64;

Query Match 7.3%; Score 89; DB 16; Length 733;
Best Local Similarity 26.6%; Pred. No. 9.8;
Matches 58; Conservative 26; Mismatches 66; Indels 68; Gaps 13;

```

```

QY 28 LNRTGDPGPGPSIQKTYDLYRLE-----HQLRSLAGTYNLVLPFPNPFDP--FNPP 77
DB 230 LNLGLHGRPGKV-----SRYLKLYFIAQDVHERASSSHYPYNRLAEAFHSDVLFRCQ 283
QY 78 RL-----GAETIPRATVDLEWRSNDKRLQNTQYAYSHLLCYLRGLNRQAATLRLR 131
DB 284 RLNLQOGKACQALARA-----TLRQPFDP-YA-----DRELALEDQA 320
QY 132 SLAHF-----CTSLQGLLGSAGVMAALGYPLPQPLPGTEPTWTGPAHSDFLQKMDDFWL 187
DB 321 SLEHLRQOSNPAMKGLRLS-LGALAANLTLDRKLAGA-----SNPDALADEQDSAL 371
QY 188 LKELQTLWRSKAD-FNRLKKMQPPAAAATLHLGAHG 224
DB 372 LDRSP-----RSLKDAFERLRQLTF-----TSLLFPHRG 400

RESULT 15
O88287 PRELIMINARY; PRT; 955 AA.
AC O88287;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE WIZS.
GN WIZ OR WIZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Matsumoto K., Ishii N., Yoshida S., Shiosaka S., Wanaka A.,
RA Toyama M.;
RT "Molecular Cloning and Distinct Developmental Expression Pattern of
RT Spliced Forms of A Novel Zinc Finger Gene wiz in The Cerebellum.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AB012266; BAA32791.1; -.
DR MGS; MGI:132363; WIZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 955 AA; 102777 MW; D4DD03A02BECCAAC CRC64;

Query Match 7.1%; Score 87; DB 11; Length 955;
Best Local Similarity 27.8%; Pred. No. 21;
Matches 35; Conservative 19; Mismatches 48; Indels 24; Gaps 7;

QY 22 LPAPVPAALNRTGDPGPGPSIQKTYDLYRLEHQLRSLAG---TYNLVLP-----PFNEPD 73
DB 615 LPLSLPLASPGKPGAGPT-----QVPR--ELSLSPITGSPSAASYLGCVATKPELQEDR 667
QY 74 FNPPLRGATLPRATVDLEWRSNDKRLQNTQYAYSHLLCYLRGL---NRQAATLRL 130
DB 668 FLPAEVKATYITQTELPFKA-KTLHEK-----TSHSSTEACELCGLYFENRKALASHAR 721
QY 131 RSLAHF 136
DB 722 AHLRQF 727

Search completed: January 27, 2003, 15:44:07
Job time : 28.5 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1963.55 Seconds
(without alignments)
11812.771 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 797

Sequence: 1 attaagcttcgcggagcc.....tctctcttcgctccccc 797

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.ste.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rtd.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	797	100.0	797	6	AR002595	Sequence
2	797	100.0	797	6	AX392086	Sequence
3	797	100.0	797	9	AF176911	Homo sapi
4	793.4	99.5	881	6	AX205024	Sequence
5	793.4	99.5	881	6	AX205042	Sequence
6	776.4	97.4	1736	9	BC012939	Homo sapi
7	749.4	94.0	1689	9	AF172854	Homo sapi
8	680	85.3	680	9	AY049779	Homo sapi
9	669.4	84.0	819	6	AR002597	Sequence
10	669.4	84.0	819	6	AX392089	Sequence
11	669.4	84.0	819	10	AF176913	Mus muscu
12	594.4	74.6	1692	6	AX205060	Sequence
13	523.4	65.7	5087	6	AR002596	Sequence
14	523.4	65.7	5087	6	AX392088	Sequence
15	523.4	65.7	5087	9	AF176912	Homo sapi
16	523.4	65.7	135116	2	AP002437	Homo sapi
17	523.4	65.7	168567	9	AP003419	Homo sapi
18	521.8	65.5	169144	2	AC005849	Homo sapi
19	429.6	53.9	269355	2	AC109138	Homo sapi
20	269.4	33.8	283	6	AX202145	Sequence
21	86.4	10.8	63347	2	AC110526	Mus muscu
22	59.4	7.5	125020	9	AF429315	Homo sapi
23	57.8	7.3	125020	9	AF429315	Homo sapi
24	47.6	6.0	648	5	AF181480	Homo sapi
25	47.6	6.0	648	5	AF181481	Plethodon
26	46.4	5.8	39552	2	AC097600	Plethodon
27	46.4	5.8	149120	2	AP005536	Rattus no
28	46.4	5.8	187727	2	AC021142	Oryza sat
29	46.2	5.8	97826	9	AL591647	Homo sapi
30	46.2	5.8	173609	9	AL451077	Human DNA
31	46	5.8	648	5	AF181482	Plethodon
32	45.6	5.7	1901	5	GGLAMB1	Chicken mRN
33	44.8	5.6	2265	9	BC009360	Homo sapi
34	44.8	5.6	16124	1	AX024384	Sequence
35	44.8	5.6	16124	6	AX024277	Sequence
36	44.8	5.6	68750	1	AF210843	Sorangium
37	44.8	5.6	68750	6	AR193029	Sequence
38	44.8	5.6	68750	6	AR199551	Sequence
39	44.8	5.6	68750	6	AR199559	Sequence
40	44.8	5.6	68750	6	AR199567	Sequence
41	44.8	5.6	68750	6	AR201097	Sequence
42	44.8	5.6	68750	6	AR208671	Sequence
43	44.4	5.6	648	5	AF181483	Plethodon
44	44	5.5	47958	2	AC091104	Homo sapi
45	44	5.5	168992	2	AC103568	Rattus no

ALIGNMENTS

RESULT 1	AR002595	Sequence 1	from patent US 5741772.	DNA	linear	PAT 04-DEC-1998
LOCUS	AR002595	Sequence 1	from patent US 5741772.	797 bp		
DEFINITION	AR002595	Sequence 1	from patent US 5741772.			
ACCESSION	AR002595	Sequence 1	from patent US 5741772.			
VERSION	AR002595.1	GI:3964149				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 797)					
AUTHORS	Chang,M.-s.					
TITLE	Neurotrophic factor NNT-1					
JOURNAL	Patent: US 5741772-A 1 21-APR-1998;					
FEATURES	Location/Qualifiers					

RESULT 2					
AX392086	LOCUS	AX392086	797 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0215977.				
ACCESSION	AX392086				
VERSION	AX392086.1	GI:19700574			
KEYWORDS	.				

RESULT 5
AX205042
LOCUS
DEFINITION
AX205042
ACCESSION
AX205042.1
VERSION
GI:15394277
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Eison, G. and Gauchat, J.F.
Sscstrfr/nnt-1 fusion protein
Patent: WO 0155219-A 1 02-AUG-2001;
PIERRE FABRE MEDICAMENT (FR)
LOCATION/Qualifiers
1. .881
/organism="Homo sapiens"
/db_xref="taxon:9606"
174. .851
/note="NNT-1"
/protein_id="CAC60178.1"
/codon_start=1
/db_xref="GI:15394278"
/translation="MDLRAGDSWGLACLTCTVHLPAVPALNRTGDPGPGSIQKTY
DLTRYLEQLRSLAGTYLNYLGPNEPDPNPRGLGAETLPRATVLEVRSNDKLR
LTQNYEAYSHLLCYLRGNROATAEELRSLAHFCTSLQGLGSIAGVMAALGYPQP
PLPTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRLKKMQPPAAAVTL
HLGAHGF"
BASE COUNT 158 a 318 c 246 g 159 t
ORIGIN
Query Match 99.5%; Score 793.4; DB 6; Length 881;
Best Local Similarity 99.9%; Pred. No. 1e-151;
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGACGCTCCGGGAGAGGAG 60
DB 85 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGACGCTCCGGGAGAGGAG 144
QY 61 CCGCACCCG 120
DB 145 CCGCACCCG 204
QY 121 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 180
DB 205 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 264
QY 181 CAGGGAGCCAGGGCGCTGGCGCTTCATCCAGAAAACCTATGACCTCACCGCTACCTGG 240
DB 265 CAGGGAGCCAGGGCGCTGGCGCTTCATCCAGAAAACCTATGACCTCACCGCTACCTGG 324
QY 241 AGCACCAACTCCGCGAGCTGGCTGGGACCTATCTGAATACCTGGGCGCCCTTTTCAAG 300
DB 325 AGCACCAACTCCGCGAGCTGGCTGGGACCTATCTGAATACCTGGGCGCCCTTTTCAAG 384
QY 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGACTCTGCCAGGGCCACTGTGTGACT 360
DB 385 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGACTCTGCCAGGGCCACTGTGTGACT 444
QY 361 TGGAGGTGTGGCGAAGCTTCATGACAAAACCTGGCGCTGACCCAGAACTACAGGGCTTACA 420
DB 445 TGGAGGTGTGGCGAAGCTTCATGACAAAACCTGGCGCTGACCCAGAACTACAGGGCTTACA 504
QY 421 GCCACCTTCTGTGTACTTGGTGGCCCTCAACCCGCTCAGGCTGCCACTGTGTGAGCTGCGCC 480
DB 505 GCCACCTTCTGTGTACTTGGTGGCCCTCAACCCGCTCAGGCTGCCACTGTGTGAGCTGCGCC 564
QY 481 GCAGCTTGGCCCACTTCTTGCAACCGCTCCAGGGCTTCAGGGGCGTTCAGGGGCGTCA 540
DB 565 GCAGCTTGGCCCACTTCTTGCAACCGCTCCAGGGCTTCAGGGGCGTTCAGGGGCGTCA 624

/note="NNT-1"
/codon_start=1
/protein_id="CAC60175.1"
/db_xref="GI:15394260"
/translation="MDLRAGDSWGLACLTCTVHLPAVPALNRTGDPGPGSIQKTY
DLTRYLEQLRSLAGTYLNYLGPNEPDPNPRGLGAETLPRATVLEVRSNDKLR
LTQNYEAYSHLLCYLRGNROATAEELRSLAHFCTSLQGLGSIAGVMAALGYPQP
PLPTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRLKKMQPPAAAVTL
HLGAHGF"
BASE COUNT 158 a 318 c 246 g 159 t
ORIGIN
Query Match 99.5%; Score 793.4; DB 6; Length 881;
Best Local Similarity 99.9%; Pred. No. 1e-151;
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGACGCTCCGGGAGAGGAG 60
DB 85 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGACGCTCCGGGAGAGGAG 144
QY 61 CCGCACCCG 120
DB 145 CCGCACCCG 204
QY 121 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 180
DB 205 TGTTAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAGCGTGTGTCAG 264
QY 181 CAGGGAGCCAGGGCGCTGGCGCTTCATCCAGAAAACCTATGACCTCACCGCTACCTGG 240
DB 265 CAGGGAGCCAGGGCGCTGGCGCTTCATCCAGAAAACCTATGACCTCACCGCTACCTGG 324
QY 241 AGCACCAACTCCGCGAGCTGGCTGGGACCTATCTGAATACCTGGGCGCCCTTTTCAAG 300
DB 325 AGCACCAACTCCGCGAGCTGGCTGGGACCTATCTGAATACCTGGGCGCCCTTTTCAAG 384
QY 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGACTCTGCCAGGGCCACTGTGTGACT 360
DB 385 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGACTCTGCCAGGGCCACTGTGTGACT 444
QY 361 TGGAGGTGTGGCGAAGCTTCATGACAAAACCTGGCGCTGACCCAGAACTACAGGGCTTACA 420
DB 445 TGGAGGTGTGGCGAAGCTTCATGACAAAACCTGGCGCTGACCCAGAACTACAGGGCTTACA 504
QY 421 GCCACCTTCTGTGTACTTGGTGGCCCTCAACCGTCAAGGCTGCCACTGTGTGAGCTGCGCC 480
DB 505 GCCACCTTCTGTGTACTTGGTGGCCCTCAACCGTCAAGGCTGCCACTGTGTGAGCTGCGCC 564
QY 481 GCAGCTGGCCCACTTCTGACACGCTCCAGGGCTGCTGGGCGAGCAVTTGGGGGCGTCA 540
DB 565 GCAGCTGGCCCACTTCTGACACGCTCCAGGGCTGCTGGGCGAGCAVTTGGGGGCGTCA 624
QY 541 TGGCAGCTCTGGGCTACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGAATC 600
DB 625 TGGCAGCTCTGGGCTACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGAATC 684
QY 601 CTGGCCCTGCCACAGTGTCTTCCAGAAAGATGACACTTCTGGCTGCTGAAGGAGC 660
DB 685 CTGGCCCTGCCACAGTGTCTTCCAGAAAGATGACACTTCTGGCTGCTGAAGGAGC 744
QY 661 TGCAGACTGGCTGTGGCGCTCGGCCCAAGGACTTCAACCGGCTCAAGAGAAAGATGACG 720
DB 745 TGCAGACTGGCTGTGGCGCTCGGCCCAAGGACTTCAACCGGCTCAAGAGAAAGATGACG 804
QY 721 CTTCAGAGCTCAGTCACTTCCAGAAAGATGACACTTCTGGCTGCTGAATCTTCT 780
DB 805 CTTCAGAGCTCAGTCACTTCCAGAAAGATGACACTTCTGGCTGCTGAATCTTCT 864
QY 781 CTCTCTGCTTCCCC 795
DB 865 CTCTCTGCTTCCCC 879

541' TGGCAGCTCTGGCTACCCACTGCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600
 |||||
 Db TGGCAGCTCTGGCTACCCACTGCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 684

601 CTGGCCCTGCCACAGTACTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGGAGC 660
 |||||
 Db CTGGCCCTGCCACAGTACTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGGAGC 744

661 TGCAGACTGCTGCTGGCTGCGCAAGGACTTCAACCGGCTCAAGAAGAGATGACGAGC 720
 |||||
 Db TGCAGACTGCTGCTGGCTGCGCAAGGACTTCAACCGGCTCAAGAAGAGATGACGAGC 804

721 CTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCT 780
 |||||
 Db CTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCT 864

781 CCTTTGCTCCCCC 795
 |||||
 Db CCTTTGCTCCCCC 879

RESULT 6
 BC012939
 LOCUS
 DEFINITION
 1736 bp mRNA linear PRI 22-AUG-2001
 Homo sapiens, Similar to cardiotrophin-like cytokine;
 neurotrophin-1/B-cell stimulating factor-3, clone MGC:21195
 IMAGE:4453813, mRNA, complete cds.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 BC012939
 MGC.
 Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1736)
 Strausberg,R.
 Direct Submission
 Submitted (20-AUG-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalon@bcm.tmc.edu
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 28 Row: b Column: 23
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6007640.
 Location/Qualifiers
 1..1736
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:21195 IMAGE:4453813"
 /tissue_type="Kidney, hypernephroma"
 /clone_lib="NIH MGC_89"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 73..750
 /codon_start=1
 /product="Similar to cardiotrophin-like cytokine;

FEATURES
 source

CDS

neurotrophin-1/B-cell stimulating factor-3"
 /protein_id="AAH12939.1"
 /db_xref="GI:15277895"

/translation="MDLRAGDSMGLACLCTVLMHLPALNRTGDPGPGSIQKTY
 DLTRYLEHQLRSLAGTYINLYGPPFNPRLGAETLPATVDLEWRLNDKLR
 LTYWEAYSHLLCYLRGLNRQAATAELRSLAHFCTSLQGLGSLAGVMAALGYPLPQ
 PLPGTEPTWTPGPHSDFLQKMDDFLLKELQTLWRSKDFNRLKKMQPAAAVTL
 HLGAHGF"

BASE COUNT 371 a 546 c 454 g 365 t
 ORIGIN

Query Match 97.4%; Score 776.4; DB 9; Length 1736;
 Best Local Similarity 99.9%; Pred. No. 2.7e-148;
 Matches 777; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 18 GCCCGGCTGCGCTTCCACTCCGCGCAGCTCCGGAGAGAGCGCACCCGGCGGCCCC 77
 |||||
 Db 1 GCCCGGCTGCGCTTCCACTCCGCGCAGCTCCGGAGAGAGCGCACCCGGCGGCCCC 60

QY 78 AGCCCCAGCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTGCTGTGC 137
 |||||
 Db 61 GGCCCCAGCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTGCTGTGC 120

QY 138 AGGTGCTCTGGCACTTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGGCT 197
 |||||
 Db 121 AGGTGCTCTGGCACTTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGGCT 180

QY 198 GGCCCTTCCATCCAGAAACCTATGACTCACCGCTACTTGGAGCAGCCAACTCCGAGC 257
 |||||
 Db 181 GGCCCTTCCATCCAGAAACCTATGACTCACCGCTACTTGGAGCAGCCAACTCCGAGC 240

QY 258 TTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCT 317
 |||||
 Db 241 TTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCT 300

QY 318 CCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTTGTGACTTTGGAGGTGTGGCGAAGC 377
 |||||
 Db 301 CCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTTGTGACTTTGGAGGTGTGGCGAAGC 360

QY 378 CTCAAATGACAACTCGGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTAC 437
 |||||
 Db 361 CTCAAATGACAACTCGGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTAC 420

QY 438 TTGGCTGGGCTCAACCGTCAAGCTGCTGAGCTGCGCGCAGCCCTGGCCACTTC 497
 |||||
 Db 421 TTGGCTGGGCTCAACCGTCAAGCTGCTGAGCTGCGCGCAGCCCTGGCCACTTC 480

QY 498 TGCACAGCTCTCAGGGCTGCTGGGAGCATTTGCGGGGTCTATGGCAGCTCTGGGGTAC 557
 |||||
 Db 481 TGCACAGCTCTCAGGGCTGCTGGGAGCATTTGCGGGGTCTATGGCAGCTCTGGGGTAC 540

QY 558 CCACTGCCCCAGCGCTGCTGGGACTGAACCACTTGGACTCTGGCCCTGGCCACAGT 617
 |||||
 Db 541 CCACTGCCCCAGCGCTGCTGGGACTGAACCACTTGGACTCTGGCCCTGGCCACAGT 600

QY 618 GACTTCTCCAGAGATGAGCACTTCTGGCTGCTGAAGGAGCTGCAGACTCGCTGTGG 677
 |||||
 Db 601 GACTTCTCCAGAGATGAGCACTTCTGGCTGCTGAAGGAGCTGCAGACTCGCTGTGG 660

QY 678 CGCTGGCCCAAGGACTTCAACCGGCTCAAGAAGAGATGAGCTTCCAGAGTGCAGTC 737
 |||||
 Db 661 CGCTGGCCCAAGGACTTCAACCGGCTCAAGAAGAGATGAGCTTCCAGAGTGCAGTC 720

QY 738 ACCCTGCACTGGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCTCC 795
 |||||
 Db 721 ACCCTGCACTGGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCTCC 778

RESULT 7
 AF172854
 LOCUS
 DEFINITION
 Homo sapiens cardiotrophin-like cytokine cDNA, complete cds.
 ACCESSION AF172854

AF172854 1689 bp mRNA linear PRI 10-SEP-1999
 DEFINITION Homo sapiens cardiotrophin-like cytokine cDNA, complete cds.
 ACCESSION AF172854

```

AF172854.1 GI:5852980
VERSION Homo sapiens.
KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
SOURCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ORGANISM 1 (bases 1 to 1689)
REFERENCE Ruben,S. and Alderson,R.F.
AUTHORS Computational EST database analysis identifies a novel member of
the neurotrophin-like cytokine family.
TITLE Biochem. Biophys. Res. Commun. 262 (1), 132-138 (1999)
JOURNAL 99382254
MEDLINE 10448081
PUBMED
REFERENCE 2 (bases 1 to 1689)
AUTHORS Shi,Y.
Direct Submission
TITLE Submitted (28-JUL-1999) Molecular Biology, Human Genome Science,
JOURNAL Inc., 9410 Key West Avenue, Rockville, MD 20850, USA
Location/Qualifiers
FEATURES
1.1689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chr="11q13.3"
/map="11q13.3"
46..723
/codon_start=1
/product="cardiotrophin-like cytokine CLC"
/protein_id="A454284.1"
/db_xref="GI:5852981"
/translation="MDLRAGDSWGLACLTVLWHLPAVPALNRTGDPGPGSIQKTY
DLTRYLHQRLSAGTYLVNLPFPNPDNPPRLGAEITLPRATVDLEWRSNDKLR
LTQNYEAYSHLLCYLRGNROAATAEURLSLAHFCTSLQGLLGSAGVMAALGYPLPQ
PLPTEPTWTPGPAHSDFLQMDDFLLKELQTLWRSKDFNRLKKKQPPAAAVTL
HLGAHGF"
BASE COUNT 349 a 530 c 448 g 362 t
ORIGIN
Query Match 94.0%; Score 749.4; DB 9; Length 1689;
Best Local Similarity 99.9%; Pred. No. 8.5e-143;
Matches 750; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 45 GCCTCCGGAGAGAGCGCAGCCCGCGCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCA 104
DB 1 GCCTCCGGAGAGAGCGCAGCCCGCGCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCCCGCCAGCA 60
QY 105 GGGGACTCGTGGGGGATGTTAGCGTCTGTGTCACCGTCTGTGGCACTTCCTGCGAGTG 164
DB 61 GGGGACTCGTGGGGGATGTTAGCGTCTGTGTCACCGTCTGTGGCACTTCCTGCGAGTG 120
QY 165 CCAGCTCTCAATCGACAGGGGACCGAGCGCTGGCGCCCTCCATCCAGAAACCTATGAC 224
DB 121 CCAGCTCTCAATCGACAGGGGACCGAGCGCTGGCGCCCTCCATCCAGAAACCTATGAC 180
QY 225 CTCACCGCTACTCGGAGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTGAC 284
DB 181 CTCACCGCTACTCGGAGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTGAC 240
QY 285 GGCCCCCCTTTCAACGAGCGAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCC 344
DB 241 GGCCCCCCTTTCAACGAGCGAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCC 300
QY 345 AGGGCCACTGTTGACTGGAGGTGGCGAAGCTCAATGACAACTGCGGCTGACCCAG 404
DB 301 AGGGCCACTGTTGACTGGAGGTGGCGAAGCTCAATGACAACTGCGGCTGACCCAG 360
QY 405 AACTACGAGGCTTACAGCCACTTCTGTGTACTTGGCTGGGCTTCAACCGTCAAGCTGCC 464
DB 361 AACTACGAGGCTTACAGCCACTTCTGTGTACTTGGCTGGGCTTCAACCGTCAAGCTGCC 420
QY 465 ACTGCTGAGTGGCGGAGCGCTGGCCCACTTCTGACCAAGCTTCCAGGGGCTGCTGGC 524
DB 421 ACTGCTGAGTGGCGGAGCGCTGGCCCACTTCTGACCAAGCTTCCAGGGGCTGCTGGC 480

```

```

QY 525 AGCATTTGGCGGCGTCAATGGCAGCTCTGGGCTACCCAGCTGCCCGCCAGCTGGCTGGGACT 584
DB 481 AGCATTTGGCGGCGTCAATGGCAGCTCTGGGCTACCCAGCTGCCCGCCAGCTGGCTGGGACT 540
QY 585 GAACCCACTTTGGACTCTTGGCCCTGCCACAGTGAATCTTCTCCAGAGATGGACGACTTC 644
DB 541 GAACCCACTTTGGACTCTTGGCCCTGCCACAGTGAATCTTCTCCAGAGATGGACGACTTC 600
QY 645 TGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCCAAGGACTTCAACCGGCTC 704
DB 601 TGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCCAAGGACTTCAACCGGCTC 660
QY 705 AAGAAAGAGATGCAGCCTCCAGCAGCTGCAGTCACTCCACCTGCAGTGGGGCTCATGGCTTC 764
DB 661 AAGAAAGAGATGCAGCCTCCAGCAGCTGCAGTCACTCCACCTGCAGTGGGGCTCATGGCTTC 720
QY 765 TGACTTCTGACCTTCTCTCTTCTGCTCCCTCC 795
DB 721 TGACTTCTGACCTTCTCTCTTCTGCTCCCTCC 751

RESULT 8
AY049779 680 bp mRNA linear PRI 23-OCT-2001
LOCUS Homo sapiens cardiotrophin-like cytokine (CLC) mRNA, complete cds.
DEFINITION
ACCESSION AY049779
VERSION AY049779.1 GI:16356642
KEYWORDS Homo sapiens.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 680)
AUTHORS Hu,X., Xu,Y., Zhang,B., Peng,X., Yuan,J. and Qiang,B.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2001) Department of Biochemistry, Institute of
Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong
Dan San Tiao, Beijing 100005, P.R. China
FEATURES
Location/Qualifiers
1..680
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..680
/genes="CLC"
3..680
/genes="CLC"
/notes="neurotrophin-1/B-cell stimulating factor-3"
/codon_start=1
/product="cardiotrophin-like cytokine"
/protein_id="A4U15436.1"
/db_xref="GI:16356643"
/translation="MDLRAGDSWGLACLTVLWHLPAVPALNRTGDPGPGSIQKTY
DLTRYLHQRLSAGTYLVNLPFPNPDNPPRLGAEITLPRATVDLEWRSNDKLR
LTQNYEAYSHLLCYLRGNROAATAEURLSLAHFCTSLQGLLGSAGVMAALGYPLPQ
PLPTEPTWTPGPAHSDFLQMDDFLLKELQTLWRSKDFNRLKKKQPPAAAVTL
HLGAHGF"
BASE COUNT 125 a 239 c 191 g 125 t
ORIGIN
Query Match 85.3%; Score 680; DB 9; Length 680;
Best Local Similarity 100.0%; Pred. No. 1.3e-128;
Matches 680; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 CCATGGAGCTCCGAGCAGGGGACTCGTGGGGATTTAGCGTCTGTGACGCTGCTCT 147
DB 1 CCATGGAGCTCCGAGCAGGGGACTCGTGGGGATTTAGCGTCTGTGACGCTGCTCT 60
QY 148 GGCACCTCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTTGGCCCTCCA 207
DB 61 GGCACCTCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTTGGCCCTCCA 120
QY 208 TCCAGAAAACCTATGACCTCACCCGCTACCTGGAGCAACCACTCCGACGCTTGGTGGGA 267

```

Db 121 TCCAGAAACCTATGACCTACCCGCTACCTGGAGACCAACTCCGACGCTTGGCTGGGA 180
 QY 268 CTTATCTGAATACCTGGGCCCCCTTTCAAGGAGCAGACTTCAACCTCCCGCGCTGG 327
 Db 181 CTTATCTGAATACCTGGGCCCCCTTTCAAGGAGCAGACTTCAACCTCCCGCGCTGG 240
 QY 328 GGGCAGAGACTCTGGCCAGGGCCACTGTTGACTTGGAGGTGGGAGGCTCAATGACA 387
 Db 241 GGGCAGAGACTCTGGCCAGGGCCACTGTTGACTTGGAGGTGGGAGGCTCAATGACA 300
 QY 388 AACTGCGGCTGACCCAGAGCTACAGAGGCTACAGCCACCTTCTGTACTTGGCTGGCC 447
 Db 301 AACTGCGGCTGACCCAGAGCTACAGAGGCTACAGCCACCTTCTGTACTTGGCTGGCC 360
 QY 448 TCAACCGTCAAGCTGCTGAGCTGCGCCGAGGCTTGGCCCACTTCTGCAACAGCC 507
 Db 361 TCAACCGTCAAGCTGCTGAGCTGCGCCGAGGCTTGGCCCACTTCTGCAACAGCC 420
 QY 508 TCCAGGGCTCTGCTGGCAGCAATGGGGGCTCATGGAGCTTGGGCTACCACTGCCCC 567
 Db 421 TCCAGGGCTCTGCTGGCAGCAATGGGGGCTCATGGAGCTTGGGCTACCACTGCCCC 480
 QY 568 AGCCGCTGCTGGGACTGAACCCACTTGGACTGCGCCGAGGCTTGGCCCACTTCTGCAACAGCC 527
 Db 481 AGCCGCTGCTGGGACTGAACCCACTTGGACTGCGCCGAGGCTTGGCCCACTTCTGCAACAGCC 540
 QY 628 AGAAGTGGAGCACTTCTGGCTGCTGAAGAGCTGCAAGCTGCTGGGCTCGGCCA 687
 Db 541 AGAAGTGGAGCACTTCTGGCTGCTGAAGAGCTGCAAGCTGCTGGGCTCGGCCA 600
 QY 688 AGGACTTCAACCCGCTCAAGAGAGATGCAAGCTTCCAGAGCTGCAAGCTGCAAGCTGCAAGCT 747
 Db 601 AGGACTTCAACCCGCTCAAGAGAGATGCAAGCTTCCAGAGCTGCAAGCTGCAAGCTGCAAGCT 660
 QY 748 TGGGGCTCATGGCTTCTGA 767
 Db 661 TGGGGCTCATGGCTTCTGA 680

RESULT 9
 AR002597
 LOCUS AR002597 819 bp DNA linear PAT 04-DEC-1998
 DEFINITION Sequence 4 from patent US 5741772.
 ACCESSION AR002597
 VERSION AR002597.1 GI:3964151
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 819)
 AUTHORS Chang, M.-s.
 TITLE Neurotrophic factor NNT-1
 JOURNAL Patent: US 5741772-A 4 21-APR-1998;
 FEATURES Location/Qualifiers
 source 1..819
 BASE COUNT 156 a 288 c 218 g 157 t
 ORIGIN

Query Match 84.0%; Score 669.4; DB 6; Length 819;
 Best Local Similarity 92.0%; Pred. No. 1.9e-126;
 Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
 QY 1 ATTAAAGCTTCGCGGAGCGGGCTCGCCCTCCCACTCGCCAGCTCGGAGAGGAG 60
 Db 5 ATTAAAGCTTCGCGGAGCGGGCTCGCCCTCCCACTCGCCAGCTTCGGAGAGGAG 64
 QY 61 CCGCACCGCGCGCGCCAG-CGCCAGCCCATGGACTCCGAGCAGGGGACTCGTGGGG 119
 Db 65 CCG 124
 QY 120 ATGTTAGCTGCTGTGACGGTGTCTGGGCACTCCCTGCTGAGTGGCAGCTCTCAATCGC 179

Db 125 ATGTTAGCTGCTGTGACCGTGTGTGGACCTCCCTGCGAGTGGCAGCTCTTAATCGC 184
 QY 180 ACAGGGAGCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 239
 Db 185 ACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 244
 QY 240 GAGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCCTTTCAAC 299
 Db 245 GAGCATCAACTCCGAGCTTGGCTGGGACCTACTCTGAACCTACTCTGGGCCCCCTTTCAAC 304
 QY 300 GAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCACTGTTGAC 359
 Db 305 GAGCCTGACTTCAATCTCTCGACTGGGGGAGAGAACTCTGCCAGGGGCACTGGTCAAC 364
 QY 360 TTGAGGCTGGGAGGAGCTCAATGACAAAATGCGGCTGACCCAGAACTACGAGGCTAC 419
 Db 365 TTGAGGCTGGGAGGAGCTCAATGACAAAATGCGGCTGACCCAGAACTATGAGGCTGAC 424
 QY 420 AGCACCTTCTGTACTTGGCTGGGCTCAACCTGAGGCTGCGCACTGCTGAGCTGGCC 479
 Db 425 AGTCACTCTGTGTTACTTGGCTGGGCTCAACCTGAGGCTGCGCACTGCTGAGCTGCA 484
 QY 480 CGCAGCTTGGGCTTCTGTGACCCAGGCTCCAGGGCTGCTGGGAGCAATTCGCGGCTG 539
 Db 485 CGTACCTTGGGCTTCTGTGACCCAGGCTCCAGGGCTGCTGGGAGCAATTCGCGGCTG 544
 QY 540 ATGGAGCTTGGGCTACCCACTGCCAGCGCTGCCAGGCTGCTGGGAGCAATTCGCGGCTG 599
 Db 545 ATGGAGCTTGGGCTACCCACTGCCAGCGCTGCCAGGCTGCTGGGAGCAATTCGCGGCTG 604
 QY 600 CTTGGGCTTGGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 659
 Db 605 CTTGGGCTTGGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 664
 QY 660 CTGAGACTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 719
 Db 665 CTGAGACTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 724
 QY 720 CTTCCAGCAGCTGAGTCACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 778
 Db 725 CTTCCAGCAGCTGAGTCACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 783

RESULT 10
 AX392089
 LOCUS AX392089 819 bp DNA linear PAT 23-MAR-2002
 DEFINITION Sequence 4 from Patent WO0215977.
 ACCESSION AX392089
 VERSION AX392089.1 GI:19700577
 KEYWORDS
 SOURCE Mus sp.
 ORGANISM Mus sp.
 REFERENCE 1
 AUTHORS Senaldi, G.
 TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 JOURNAL Methods and compositions for treating ige-related disease using nn t-1 inhibitors
 Patent: WO 0215977-A 4 28-FEB-2002;
 Amgen Inc. (US)
 FEATURES Location/Qualifiers
 source 1..819
 CDS
 /organism="Mus sp."
 /db_xref="taxon:10095"
 95..772
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD28986.1"
 /db_xref="GI:19700578"
 /translation="MDLRAGDSWGLACLCTVHLPAVPALNRTGDPGPGPSIOKTY
 DLTRYLHQRLSLAGTYLNYLGPFPNPFDPFNPRLGAEETLPRTATVNVKRSNDRLR
 LTQNEYASHLLCYLRGLNRQAELRRSLAHFCTSLQGLLSIAGVMATLGYPLPQ

Mus musculus.		Mus musculus.	
Mus musculus		Mus musculus	
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 819)		1 (bases 1 to 819)	
Senaldi, G., Varnum, B.C., Sarmiento, U., Starnes, C., Lile, J.,		Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C.,	
Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Shaklee, C.,		Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Shaklee, C.,	
Manu, F., Simonet, S., Boone, T. and Chang, M.-S.		Manu, F., Simonet, S., Boone, T. and Chang, M.-S.	
Freeman, D., Manu, F., Simonet, W.S., Boone, T. and Chang, M.S.		Freeman, D., Manu, F., Simonet, W.S., Boone, T. and Chang, M.S.	
Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the		Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the	
IL-6 family		IL-6 family	
Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)		Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)	
9432254		9432254	
10500198		10500198	
2 (bases 1 to 819)		2 (bases 1 to 819)	
Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C.,		Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C.,	
Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Shaklee, C.,		Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Shaklee, C.,	
Manu, F., Simonet, S., Boone, T. and Chang, M.-S.		Manu, F., Simonet, S., Boone, T. and Chang, M.-S.	
Direct Submission		Direct Submission	
Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive,		Submitted (11-AUG-1999) Amgen, Inc., One Amgen Center Drive,	
Thousand Oaks, CA 91320, USA		Thousand Oaks, CA 91320, USA	
Location/Qualifiers		Location/Qualifiers	
1. 819		1. 819	
/organism="Mus musculus"		/organism="Mus musculus"	
/db_xref="taxon:10090"		/db_xref="taxon:10090"	
95.1772		95.1772	
/codon_start=1		/codon_start=1	
/product="neurotrophin-1/B-cell stimulating factor-3"		/product="neurotrophin-1/B-cell stimulating factor-3"	
/protein_id="AAF00993.1"		/protein_id="AAF00993.1"	
/db_xref="GI:6007645"		/db_xref="GI:6007645"	
/translation="MDLRAGDSWGLACLTVLWHLPAVPAALNRTGPPGSPISQKTY		/translation="MDLRAGDSWGLACLTVLWHLPAVPAALNRTGPPGSPISQKTY	
DUTRYEHLRLSLAGTYLNYLPPNEPDPNPRLGAETLPRAVLNLEWVRSNDRLR		DUTRYEHLRLSLAGTYLNYLPPNEPDPNPRLGAETLPRAVLNLEWVRSNDRLR	
LTONYEAYSHLLCYLRGNRQATAEELRSLAHFCTSLQGLGSIAGVMATLGYPLPO		LTONYEAYSHLLCYLRGNRQATAEELRSLAHFCTSLQGLGSIAGVMATLGYPLPO	
PLPGETPAWAPGPAHSDFLQKMDDFLLKELQTLWLRSAKDFNRLKKMQPPAASVTL		PLPGETPAWAPGPAHSDFLQKMDDFLLKELQTLWLRSAKDFNRLKKMQPPAASVTL	
HLEAHGF"		HLEAHGF"	
BASE COUNT	156 a 288 c 218 g 157 t	BASE COUNT	156 a 288 c 218 g 157 t
ORIGIN		ORIGIN	
Query Match 84.0%; Score 669.4; DB 10; Length 819;		Query Match 84.0%; Score 669.4; DB 10; Length 819;	
Best Local Similarity 92.0%; Pred. No. 1.9e-126;		Best Local Similarity 92.0%; Pred. No. 1.9e-126;	
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;		Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;	
QY	1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACCTCCCGCCAGCTCCCGGAGAGGAG 60	QY	1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACCTCCCGCCAGCTCCCGGAGAGGAG 60
DB	5 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACCTCCCGCCAGCTTCGGGAGGAG 64	DB	5 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACCTCCCGCCAGCTTCGGGAGGAG 64
QY	61 CCGCAGCCCGCGCGCCAG-CCCCAGCCCCATGACCTCCGAGAGGGGACTCGTGGGG 119	QY	61 CCGCAGCCCGCGCGCCAG-CCCCAGCCCCATGACCTCCGAGAGGGGACTCGTGGGG 119
DB	65 CCGCGCCCGCGCGCCAGCCCCCGCCAGCCCCATGACCTCCGAGAGGGGACTCGTGGGG 124	DB	65 CCGCGCCCGCGCGCCAGCCCCCGCCAGCCCCATGACCTCCGAGAGGGGACTCGTGGGG 124
QY	120 ATGTTAGCTTCGCTGTCACCGGTCTTGGCACCTCCCTGCGAGTCCAGTCCCAATCGC 179	QY	120 ATGTTAGCTTCGCTGTCACCGGTCTTGGCACCTCCCTGCGAGTCCAGTCCCAATCGC 179
DB	125 ATGTTAGCTTCGCTGTCACCGGTCTTGGCACCTCCCTGCGAGTCCAGTCCCAATCGC 184	DB	125 ATGTTAGCTTCGCTGTCACCGGTCTTGGCACCTCCCTGCGAGTCCAGTCCCAATCGC 184
QY	180 ACAGGAGACCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 239	QY	180 ACAGGAGACCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 239
DB	185 ACAGGAGATCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 244	DB	185 ACAGGAGATCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTG 244
QY	240 GAGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAATACCTGGGCCCCCTTTCAAC 299	QY	240 GAGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAATACCTGGGCCCCCTTTCAAC 299
DB	245 GAGCATCAACTCCGAGCTTAGCTGGGACCTACCTGAACCTACCTGGGCCCCCTTTCAAC 304	DB	245 GAGCATCAACTCCGAGCTTAGCTGGGACCTACCTGAACCTACCTGGGCCCCCTTTCAAC 304
QY	300 GAGCCAGACTTCAACCTCCCGCTCGCGCGCGAGAGACTCTGCCAGGGGCACTGTGTAC 359	QY	300 GAGCCAGACTTCAACCTCCCGCTCGCGCGCGAGAGACTCTGCCAGGGGCACTGTGTAC 359
DB	305 GAGCCTGACTTCAATCTCTCCGCTGGGGGCGAGAACTCTGCCAGGGGCACTGTGTAC 364	DB	305 GAGCCTGACTTCAATCTCTCCGCTGGGGGCGAGAACTCTGCCAGGGGCACTGTGTAC 364
QY	360 TTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACAGAGGCTAC 419	QY	360 TTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACAGAGGCTAC 419
DB	365 TTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACAGAGGCTAC 424	DB	365 TTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACAGAGGCTAC 424
QY	420 AGCCACCTCTGCTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 479	QY	420 AGCCACCTCTGCTGTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 479
DB	425 AGTCACTCTCTGTTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 484	DB	425 AGTCACTCTCTGTTACTTGGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCCG 484
QY	480 CCGAGCTGGCCCACTTCTGACACCGCTCCAGGCGCTGCTGGGCGAGCAATTCGGGGCTC 539	QY	480 CCGAGCTGGCCCACTTCTGACACCGCTCCAGGCGCTGCTGGGCGAGCAATTCGGGGCTC 539
DB	485 CGTAGCTGGCCCACTTCTGACACCGCTCCAGGCGCTGCTGGGCGAGCAATTCGGGGCTC 544	DB	485 CGTAGCTGGCCCACTTCTGACACCGCTCCAGGCGCTGCTGGGCGAGCAATTCGGGGCTC 544
QY	540 ATGGCAGCTCTGGCTACCCACTGCCCCGCGCTGCTGGGACTGAACCACTTGGACT 599	QY	540 ATGGCAGCTCTGGCTACCCACTGCCCCGCGCTGCTGGGACTGAACCACTTGGACT 599
DB	545 ATGGCAGCTCTGGCTACCCACTGCCCCGCGCTGCTGGGACTGAACCACTTGGACT 604	DB	545 ATGGCAGCTCTGGCTACCCACTGCCCCGCGCTGCTGGGACTGAACCACTTGGACT 604
QY	600 CTGGCCCTGCCACAGTGAATCTCTCCAGAAATGGAAGCACTTCTGGCTGCTGAAGGAG 659	QY	600 CTGGCCCTGCCACAGTGAATCTCTCCAGAAATGGAAGCACTTCTGGCTGCTGAAGGAG 659
DB	605 CTGGCCCTGCCACAGTGAATCTCTCCAGAAATGGAAGCACTTCTGGCTGCTGAAGGAG 664	DB	605 CTGGCCCTGCCACAGTGAATCTCTCCAGAAATGGAAGCACTTCTGGCTGCTGAAGGAG 664
QY	660 CTGACAGCTGCTGGGCGCTCGCCAGCACTTCAACCGCTCAGAGAGAGATGAG 719	QY	660 CTGACAGCTGCTGGGCGCTCGCCAGCACTTCAACCGCTCAGAGAGAGATGAG 719
DB	665 CTGACAGCTGCTGGGCGCTCGCCAGCACTTCAACCGCTCAGAGAGAGATGAG 724	DB	665 CTGACAGCTGCTGGGCGCTCGCCAGCACTTCAACCGCTCAGAGAGAGATGAG 724
QY	720 CTTCAGGAGCTGAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTT 778	QY	720 CTTCAGGAGCTGAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTT 778
DB	725 CTTCAGGAGCTGAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTT 783	DB	725 CTTCAGGAGCTGAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTT 783
RESULT 11		RESULT 11	
AF176913		AF176913	
LOCUS	AF176913	LOCUS	AF176913
DEFINITION	Mus musculus neurotrophin-1/B-cell stimulating factor-3 mRNA, complete cds.	DEFINITION	Mus musculus neurotrophin-1/B-cell stimulating factor-3 mRNA, complete cds.
ACCESSION	AF176913	ACCESSION	AF176913
VERSION	AF176913.1	VERSION	AF176913.1
KEYWORDS	GI:6007644	KEYWORDS	GI:6007644

Db 3423 CAGAGACTCTGCCCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482
Qy 391 TCGGGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTTACTTGGCTGGCCTCA 450
Db 3483 TCGGGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTTACTTGGCTGGCCTCA 3542
Qy 451 ACCGTCAGGCTGCCACTGCTGAGCTGCGCGGAGCTGGCCCCACTTCTGCACCAGCCTCC 510
Db 3543 ACCGTCAGGCTGCCACTGCTGAGCTGCGCGGAGCTGGCCCCACTTCTGCACCAGCCTCC 3602
Qy 511 AGGGCTCTGCTGGGCGAGCATTTGGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 570
Db 3603 AGGGCTCTGCTGGGCGAGCATTTGGGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
Qy 571 CGCTGCTGGGACTGAAACCCACTTGGACTCTGGGCCCTGCCACAGTGAATTCCTCCAGA 630
Db 3663 CGCTGCTGGGACTGAAACCCACTTGGACTCTGGGCCCTGCCACAGTGAATTCCTCCAGA 3722
Qy 631 AGATGGAGGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG 690
Db 3723 AGATGGAGGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG 3782
Qy 691 ACTTCAACCGGCTCAAGAAAGATGCGAGCCTCCAGCAGCTGCAGTCAACCTGCACCTGG 750
Db 3783 ACTTCAACCGGCTCAAGAAAGATGCGAGCCTCCAGCAGCTGCAGTCAACCTGCACCTGG 3842
Qy 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTGCTCTTCTGCTCCCCC 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTGCTCTTCTGCTCCCCC 3887

Search completed: February 1, 2003, 17:34:10
Job time : 1980.55 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1618.85 Seconds
(without alignments)
7973.437 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 797

Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgccccccc 797

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	740.4	92.9	887	14	BQ948158
2	656.4	82.4	1157	14	BQ940483
3	641.6	80.5	853	12	BG437538
4	621	77.9	1053	12	BG164929
5	590.8	74.1	594	14	BM763333
6	577.4	72.4	580	14	BM848189

7	571.4	71.7	573	14	BM840863
8	563	70.6	955	9	AL543945
9	528.8	66.3	532	14	BM846370
10	520.4	65.3	522	14	BM841897
11	518.4	65.0	542	14	BM821005
12	514.8	64.6	529	14	BM847924
13	458.8	57.6	569	14	BM846748
14	412	51.7	476	10	BE632644
15	405.4	50.9	420	14	BM764238
16	403.4	50.6	458	13	BM363136
17	387	48.6	407	9	AI752561
18	371.4	46.6	482	14	BM846622
19	277.6	34.8	913	13	BI912197
20	274.4	34.4	512	9	AI390475
21	227.2	28.5	488	9	AA015243
22	227	28.5	291	10	BB864730
23	226	28.4	915	12	BF035982
24	174.4	21.9	552	12	BG095271
25	165.2	20.7	440	12	BG148676
26	161	20.2	691	12	BF213570
27	150.2	20.1	560	9	AL800340
28	116	14.6	655	13	BJ096761
29	114.6	14.4	633	13	BJ099801
30	63.4	8.0	711	12	BG758365
31	57	7.2	925	17	CNS0091P
32	56.8	7.1	180	9	AA204015
33	52.2	6.5	925	17	CNS0091P
34	50.2	6.3	1023	17	AG128304
35	49.6	6.2	816	17	CNS037NP
36	48.4	6.1	935	17	AG126234
37	48.4	6.1	1041	14	BQ058321
38	48.4	6.1	1469	13	BM552519
39	48.2	6.0	1090	17	AG072433
40	48	6.0	958	9	AL547011
41	47.2	5.9	927	14	BQ680645
42	47.2	5.9	1034	14	BQ641825
43	47	5.9	1651	12	BG809816
44	46.8	5.9	941	14	BQ954342
45	46.8	5.9	955	14	BQ949322

ALIGNMENTS

RESULT 1
BQ948158
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ948158
AGENCY: 8813192 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:6428214
5', mRNA sequence.
BQ948158
EST.
BQ948158.1 GI:22363636
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 887)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing By: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM2614 row: h column: 07
High quality sequence stop: 674.
Location/Qualifiers
1. 887

FEATURES
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6428214"
/tissue_lib="NIH MGC_101"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 154 a 325 c 240 g 167 t 1 others
ORIGIN
Query Match 92.9%; Score 740.4; DB 14; Length 887;
Best Local Similarity 99.7%; Pred. No. 5.9e-151;
Matches 752; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATTAAAGCTTGGCGGAGCGCGCTCGCCCTCCACTCCGACGCTCCCGGAGAGGAG 60
DB 8 ATTAAAGCTTGGCGGAGCGCGCTCGCCCTCCACTCCGACGCTCCCGGAGAGGAG 67
QY 61 CCGCACCCGCGCGCCAGCCGCCATGAGCTCCGAGCAGGGGACTCGTGGGGA 120
DB 68 CCGCACCCGCGCGCCAGCCGCCATGAGCTCCGAGCAGGGGACTCGTGGGGA 127
QY 121 TGTTAGCGTCTGTGACGGTGTCTGGACCTCCCTCGAGTGCAGCTCTCAATCGCA 180
DB 128 TGTTAGCGTCTGTGACGGTGTCTGGACCTCCCTCGAGTGCAGCTCTCAATCGCA 187
QY 181 CAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTACCCGCTACCTGG 240
DB 188 CAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTACCCGCTACCTGG 247
QY 241 AGCACCACTCCGAGCTTGGCTGGAGCTATCTGAATACCTGGGCCCCCTTTCAAG 300
DB 248 AGCACCACTCCGAGCTTGGCTGGAGCTATCTGAATACCTGGGCCCCCTTTCAAG 307
QY 301 AGCCAGACTTCAACCTCCCGCC--TGGGGGAGAGACTCTGCCAGGGGCACTGTGTAC 359
DB 308 AGCCAGACTTCAACCTCCCGCCATGGGGGAGAGACTCTGCCAGGGGCACTGTGTAC 367
QY 360 TTGGAGGTGTGGCGAAGCCTCAATGACAAACTCGGCTGACCCAGAACTACGAGGCTAC 419
DB 368 TTGGAGGTGTGGCGAAGCCTCAATGACAAACTCGGCTGACCCAGAACTACGAGGCTAC 427
QY 420 AGCCACCTTCTGTGTTACTTGGTGGCCCTCAACCGTCAGGCTGCCACTGTGAGCTGCGC 479
DB 428 AGCCACCTTCTGTGTTACTTGGTGGCCCTCAACCGTCAGGCTGCCACTGTGAGCTGCGC 487
QY 480 CGCAGCTGGCCCACTTCTGACACAGCTCCAGGGCTGTGGGCGAGCAATTGGGGGCTC 539
DB 488 CGCAGCTGGCCCACTTCTGACACAGCTCCAGGGCTGTGGGCGAGCAATTGGGGGCTC 547
QY 540 ATGGGAGCTCTGGGCTACCCACTGCCCCAGCCGCTGGCTGGGACTGAACCACTTGACT 599
DB 548 ATGGGAGCTCTGGGCTACCCACTGCCCCAGCCGCTGGCTGGGACTGAACCACTTGACT 607
QY 600 CTGGGCTGGCCACAGTGAATCTCTCCAGAGATGACGACTTCTGGCTGTGAAGGAG 659
DB 608 CTGGGCTGGCCACAGTGAATCTCTCCAGAGATGACGACTTCTGGCTGTGAAGGAG 667
QY 660 CTGACAGCTGGCTGTGGGCTCGGGCAAGGACTTCAACCGGCTCAAGAGAAATGAG 719
DB 668 CTGACAGCTGGCTGTGGGCTCGGGCAAGGACTTCAACCGGCTCAAGAGAAATGAG 727
QY 720 CTTCAGAGAGCTGAGTACCTGCACTGGGG 753
DB 728 CTTCAGAGAGCTGAGTACCTGCACTGGGG 761
```

```
RESULT 2
BQ940483
LOCUS
DEFINITION
BQ940483 1157 bp mRNA linear EST 21-AUG-2002
AGENCOURT 8864294 Lupski_sciatic_nerve Homo sapiens cDNA clone
IMAGE:6197786 5', mRNA sequence.
ACCESSION
BQ940483
VERSION
BQ940483.1 GI:22355961
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1157)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13608 row: g column: 03
High quality sequence stop: 572.
location/Qualifiers
1..1157
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6197786"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TGACCCAGCGCTCCG-3' and
5'-GACTAGTTCTAGATCGCGCTGCTGCGGCTCTGCGACCTCCCTGCACT 163
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 221 a 374 c 346 g 216 t
ORIGIN
Query Match 82.4%; Score 656.4; DB 14; Length 1157;
Best Local Similarity 97.4%; Pred. No. 1.2e-132;
Matches 678; Conservative 0; Mismatches 16; Indels 2; Gaps 1;
QY 104 AGGGGACTCGTGGGGGATGTTAGCGTGTGCGGCTGCTGCGGCTCTGCGACCTCCCTGCACT 163
DB 58 AGGGGACTCGTGGGGGATGTTAGCGTGTGCGGCTGCTGCGGCTCTGCGACCTCCCTGCACT 117
QY 164 GCCAGCTCTCAATCGCAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGA 223
DB 118 GCCAGCTCTCAATCGCAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGA 177
QY 224 CCTCACCCGCTTACCTGGAGCACCACCTCGCAGCTTGGCTGGGACCTATCTGAACCTACCT 283
DB 178 CCTCACCCGCTTACCTGGAGCACCACCTCGCAGCTTGGCTGGGACCTATCTGAACCTACCT 237
QY 284 GGGCCCCCTTTACAGAGCCAGACTTCAACCTCCCGGCTGGGGGAGAGACTCTGCC 343
DB 238 GGGCCCCCTTTACAGAGCCAGACTTCAACCTCCCGGCTGGGGGAGAGACTCTGCC 297
QY 344 CAGGGCCACTGTGACTTGGAGGTGGCGAAGCTCAATGACAACTCGGCTGACCCA 403
```


Db 542 CTACCCACTGCCCCCAGCGCTGCTGGAGTGAACCCACTTTGGACTCTCTGGCCCTTGC 601
Qy 612 CACAGTGAATCTCTCCAGAGATGAGACACTTCTGGCTGCTGAAGAGAGCTGCAGACCTGG 671
Db 602 CACAGTGAATCTCTCCAGAGATGAGACACTTCTGGCTGCTGAAGAGAGCTGCAGACCTGG 661
Qy 672 CTGTGGCGCTCGGCCCAAGGACTTCAACCGGCTCAAGAGAAGAGATCCAGCTCCAGCAGCT 731
Db 662 CTGTGGCGCTCGGCC-AGGACTTCCACCGTCCAGACGAAGATTGCAGCTCCAGCAGTT 720
Qy 732 GCAGTCAACCTTCACCTCGCGGCTCATGGCTTCTGACTTCTGACTT 778
Db 721 GCAGTCAACCTTCACCTCGCGGCTCATGGCTTCTGACTTCTGACTT 766

RESULT 5
BM763333
LOCUS K-EST0044645 S13KMS5 Homo sapiens cDNA clone S13KMS5-18-E09 5',
DEFINITION mRNA sequence.
ACCESSION BM763333
VERSION BM763333.1 GI:19092948
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 18 row: E column: 09
High quality sequence stop: 594.
Location/Qualifiers

FEATURES
source
1..594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S13KMS5-18-E09"
/clone_lib="S13KMS5"
/tissue_type="myeloma"
/cell_line="KMS-5"
/lab_host="Top10F"
/note="vector: pcNS; Site_1: EcoRI; Site_2: NotI; The poly
(A) + RNA was dephosphorylated with bacterial alkaline
phosphatase (BAP) and then deapped with tabacco acid
pyrophosphatase (TAP). The deapped intact mRNA was
ligated with DNA-RNA linker including EcoR I site by
treatment of T4 RNA ligase and the first strand cDNA was
synthesized from oligo dt-selected mRNA by priming with
dt-tailed vector. The cDNA vector was circularized with E.
coli DNA ligase after digestion of EcoRI which site is
also included in vector. An RNA strand converted to a DNA
strand by Okayama-Berg method. The obtained cDNA vectors
were used for transformation of competent cells E. coli
Top10F, by electroporation method. The cDNA libraries
constructed by this method are full-length enriched cDNA
library."

BASE COUNT 101 a 226 c 164 g 103 t
ORIGIN
Query Match 74.1%; Score 590.8; DB 14; Length 594;
Best Local Similarity 99.7%; Pred. No. 1.8e-118;

1 (bases 1 to 1053)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0244 row: m column: 14
High quality sequence start: 3
High quality sequence stop: 675.
Location/Qualifiers

1..1053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4453813"
/clone_lib="NIH_MGC_89"
/tissue_type="hypertrophoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 210 a 351 c 292 g 200 t
ORIGIN

Query Match 77.9%; Score 621; DB 12; Length 1053;
Best Local Similarity 94.1%; Pred. No. 5.9e-125;
Matches 722; Conservative 0; Mismatches 35; Indels 10; Gaps 7;
Qy 20 CGCGGCTCGCCTCCATCCGAGCTCCGGAGAGAGCGGCGCGCGCGCCAG 79
Db 2 CGCGGCTCGCCTCCATCCGAGCTCCGGAGAGAGCGGCGCGCGCGCCAG 61
Qy 80 CCCAG-CCCCATGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGGTGCTGTGA 138
Db 62 CCCCAGTCCCCATGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGGTGCTGTGA 121
Qy 139 CGGTGCTTGCCAGCTCCCTGAGTCCAGCTTCAATCGCAGGGGACCCAGGGCCTG 198
Db 122 CGGTGCTTGCCAGCTCCCTGAGTCCAGCTTCAATCGCAGGGGACCCAGGGCCTG 181
Qy 199 GCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCT 258
Db 182 GCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCT 241
Qy 259 TGGCTGGGACCTATCTGAATCTCTGGGCCCCC-----TTTCAACGAGCAGACTTCAAC 314
Db 242 TGGCTGGGACCTATCTGAATCTCTGGGCCCCC-----TTTCAACGAGCAGACTTCAAC 301
Qy 315 CTCTCCCGCTGGGGGAGAGACTCTG-CCGAGGCGACTGTTGACTTGGAGGTGTGGCG 373
Db 302 CTCTCCCGCTGGGGGAGAGACTCTG-CCGAGGCGACTGTTGACTTGGAGGTGTGGCG 361
Qy 374 AGCCCTCAATGACAACTCGCGGCTGACCCAGAACTACGAGCGCTACAGCCACTTCTGTG 433
Db 362 AGCCCTCAATGACAACTCGCGGCTGACCCAGAACTACGAGCGCTACAGCCACTTCTGTG 421
Qy 434 TTACTTGGCTGGCTCAACCGTCAAGCTGCGACTGTGAGTGGCGCGCGAGCTGGCCCA 493
Db 422 TTACTTGGCTGGCTCAACCGTCAAGCTGCGACTGTGAGTGGCGCGCGAGCTGGCCCA 481
Qy 494 CTCTGACACGAGCTCCAGGCGCTGCTGGGAGCATTCGCGGCGCTCATGGCAGCTCTGGG 553
Db 482 GTTCTGACACGAGCTCCAGGCGCTGCTGGGAGCATTCGCGGCGCTCATGGCAGCTCTGGG 541
Qy 554 CTACCCACTG-CCCCAGCGCTGCTGGGAGCTGAACCCAC-TTGGACTCTCTGGCCCTGCC 611

[illegible]

```

/db_xref="taxon:9606"
/clone="S13KMS5-48-H08"
/clone_lib="S13KMS5"
/tissue_type="myeloma"
/cell_line="KMS-5"
/lab_host="Top10F,"
/note="Vector; PCNS; Site 1: EcoRI; Site 2: NotI; The poly
(A) + RNA was dephosphorylated with bacterial alkaline
phosphatase (BAP) and then decapped with tobacco acid
pyrophosphatase (TAP). The decapped intact mRNA was
ligated with DNA-RNA linker including EcoR I site by
treatment of T4 RNA ligase and the first strand cDNA was
synthesized from oligo dt-selected mRNA by priming with
dt-tailed vector. The dt-tailed vector was adjusted to
have about 60nt. The cDNA vector was circularized with E.
coli DNA ligase after digestion of EcoRI which site is
also included in vector. An RNA strand converted to a DNA
strand by Okayama-Berg method. The obtained cDNA vectors
were used for transformation of competent cells E. coli
top10F by electroporation method. The cDNA libraries
constructed by this method are full-length enriched cDNA
library."
98 a 220 c 162 g 100 t

```

BASE COUNT	98 a	220 c	162 g	100 t	ORIGIN
Query Match	72.4%	Score 577.4	DB 14	Length 580	
Best Local Similarity	99.8%	Pred. No. 1.5e-115			
Matches 578	Conservative 0	Mismatches 1	Indels 0	Gaps 0	
QY	31	CTCCCACTCGCCAGAGCTCGGAGAGGAGCGCACC	CGGCGCGGCGCCAGCCCGCCAGCCCA	90	
Db	2	CTCCCACTCGCCAGAGCTCGGAGAGGAGCGCACC	CGGCGCGGCGCGGCGCCAGCCCGCCAGCCCA	61	
QY	91	TGACACTTCGAGCAGGGGACTCGTGGGGAGATTAG	CGGTGCTGTGCACGGTGCCTCTGGC	150	
Db	62	TGACACTTCGAGCAGGGGACTCGTGGGGAGATTAG	CGGTGCTGTGCACGGTGCCTCTGGC	121	
QY	151	ACCTCCCTGCAGTGCACGCTCTCAATCCACAGG	GGAGCCACAGGGCTGGCCCTCCATCC	210	
Db	122	ACCTCCCTGCAGTGCACGCTCTCAATCCACAGG	GGAGCCACAGGGCTGGCCCTCCATCC	181	
QY	211	AGAAACCTATGACTCTACCCGCTACCTGGAGCA	CCAACTCCGACAGCTTGGCTGGGACCT	270	
Db	182	AGAAACCTATGACTCTACCCGCTACCTGGAGCA	CCAACTCCGACAGCTTGGCTGGGACCT	241	
QY	271	ATCTGAATCTACTGGGCGCCCTTTCAAGAGCC	AGACTTCAACCTCCCGCTGGGG	330	
Db	242	ATCTGAATCTACTGGGCGCCCTTTCAAGAGCC	AGACTTCAACCTCCCGCTGGGG	301	
QY	331	CAGAGACTCTGCCAGGGCCACTGTGTACTTGA	GTGGAGGTGGCGAAGGCTCAATGACAAC	390	
Db	302	CAGAGACTCTGCCAGGGCCACTGTGTACTTGA	GTGGAGGTGGCGAAGGCTCAATGACAAC	361	
QY	391	TGCGGCTGACCCAGAACTACAGGCGCTACAG	CCACTTCTGTACTTGGCTGGGCTCA	450	
Db	362	TGCGGCTGACCCAGAACTACAGGCGCTACAG	CCACTTCTGTACTTGGCTGGGCTCA	421	
QY	451	ACCGTCAGGCTGCCACTTGCTAGCTGGCGC	GCAGCTTGGCCCACTTGTGACAGCGCTCC	510	
Db	422	ACCGTCAGGCTGCCACTTGCTAGCTGGCGC	GCAGCTTGGCCCACTTGTGACAGCGCTCC	481	
QY	511	AGGCGCTGCTGGGAGCAATTGCGGGCGCTAT	GGCAGCTCTGGGCTACCCACTGCCCGAGC	570	
Db	482	AGGCGCTGCTGGGAGCAATTGCGGGCGCTAT	GGCAGCTCTGGGCTACCCACTGCCCGAGC	541	
QY	571	CGCTGCTGGGACTGAACCCACTTGGACTCCT	GGCCCTG	609	
Db	542	CGCTGCTGGGACTGAACCCACTTGGACTCCT	GGCCCTG	580	

RESULT 7	BM840863	573 bp	mRNA	linear	EST 06-MAR-2002
BM840863					
LOC					

GENUS	BM841897	522 bp	mRNA	linear	EST 06-MAR-2002
DEFINITION	K-EST0119216 S13KMS5 Homo sapiens cDNA clone S13KMS5-34-A05 5', mRNA sequence.				
FEATURES	BM841897				
ORIGIN	BM841897.1 GI:19198306 human.				
ORIGIN	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 522) Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.				
TITLE	21C Frontier Korean EST Project 2001				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 34 row: A column: 05 High quality sequence stop: 522. Location/Qualifiers 1..522 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="S13KMS5-34-A05" /clone_lib="S13KMS5" /tissue_type="myeloma" /cell_line="KMS-5" /db_host="Top10P" /note=vector: pCNS; Site 1: EcoRI; Site 2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."				
FEATURES	source				
BASE COUNT	90 a 195 c 148 g 89 t				
ORIGIN					
Query Match	65.3%;	Score 520.4;	DB 14;	Length 522;	
Best Local Similarity	99.8%;	Pred. No. 3.6e-103;			
Matches 521;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
31	CTCCCACTCCGCGAGCCTTCGGGAGAGGACCGCACC	CGCGCGCGGCCCGCCAGCCCCAGCCCCA	90		
1	CTCCCACTCGCGAGCCTTCGGGAGAGGACCGCACC	CGCGCGCGGCCCGCCAGCCCCAGCCCCA	60		
91	TGGACCTTCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCCTGTGCACGGTGTCTCTGGC	150			
61	TGGACCTTCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCCTGTGCACGGTGTCTCTGGC	120			
151	ACCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCGCTGGCCCTTCATCC	210			
121	ACCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCGCTGGCCCTTCATCC	180			
211	AGAAACCTATGACCTTACCGGCTACTCGGAGCACCACCTCCGACGTTGGCTGGGACCT	270			
181	AGAAACCTATGACCTTACCGGCTACTCGGAGCACCACCTCCGACGTTGGCTGGGACCT	240			

ORIGIN

Query Match 65.0%; Score 518.4; DB 14; Length 542;
 Best Local Similarity 98.2%; Pred. No. 1e-102;
 Matches 535; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 9 TTCGCGGAGCGCGGCTCGCCCTCCCACTCGGCGAGGAGGAGCGCACCC 68
 Db 1 TTCGCGGAGCGCGGCTCGCCCTCCCACTCGGCGAGGAGGAGCGCACCC 57
 QY 69 GCGCGGCGGAGCGCGGCTCGCCCTCCCACTCGGCGAGGAGGAGGAGGAGG 128
 Db 58 GCGCGGCGGAGCGCGGCTCGCCCTCCCACTCGGCGAGGAGGAGGAGGAGG 117
 QY 129 TGCCTGTGACGAGGCTGTCGACACTCGGCGAGGAGGAGGAGGAGGAGG 188
 Db 118 TGCCTGTGACGAGGCTGTCGACACTCGGCGAGGAGGAGGAGGAGGAGG 177
 QY 189 CAGGGGCTGCGGCTCGCCCTCCCACTCGGCGAGGAGGAGGAGGAGGAGG 248
 Db 178 CAGGGGCTGCGGCTCGCCCTCCCACTCGGCGAGGAGGAGGAGGAGGAGG 237
 QY 249 CTCGCGAGCTTGGCTGGGAGCTATCTGAACCTACCTGGGCGGCGGCGG 308
 Db 238 CTCGCGAGCTTGGCTGGGAGCTATCTGAACCTACCTGGGCGGCGGCGG 297
 QY 309 TTCAACCCCTCCCGCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 368
 Db 298 TTCAACCCCTCCCGCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 357
 QY 369 TGGCGAAGCTTCAATGACAACTCGGCTGACCCAGAGGAGGAGGAGGAGG 428
 Db 358 TGGCGAAGCTTCAATGACAACTCGGCTGACCCAGAGGAGGAGGAGGAGG 417
 QY 429 CTGTGTACTTGGCTGGGCTCAACCGTCAAGCTGAGCTGAGCTGCGCGG 488
 Db 418 CTGTGTACTTGGCTGGGCTCAACCGTCAAGCTGAGCTGAGCTGCGCGG 477
 QY 489 GCCACTTCTGACAGGCTTCCAGGCGCTGCTGGGAGGAGGAGGAGGAGG 548
 Db 478 GCCACTTCTGACAGGCTTCCAGGCGCTGCTGGGAGGAGGAGGAGGAGG 537
 QY 549 CTGGG 553
 Db 538 CTGGG 542

RESULT 12

BM847924
 LOCUS
 DEFINITION K-EST0127535 S13KMS5 Homo sapiens cDNA clone S13KMS5-57-A03 5',
 mRNA sequence.
 ACCESSION BM847924
 VERSION BM847924.1 GI:19204323
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 529)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 57 row: A column: 03

FEATURES

High quality sequence stop: 529.
 Location/Qualifiers
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="S13KMS5-57-A03"
 /clone_lib="S13KMS5"
 /tissue_type="myeloma"
 /cell_line="KMS-5"
 /lab_host="Top10F"
 /notes="Vector: pcNS; Site 1: EcoRI; Site 2: NotI; The poly
 (A)+ RNA was dephosphorylated with bacterial alkaline
 phosphatase (BAP) and then decapped with tobacco acid
 pyrophosphatase (TAP). The decapped intact mRNA was
 ligated with DNA-RNA linker including EcoR I site by
 treatment of T4 RNA ligase and the first strand cDNA was
 synthesized from oligo dT-selected mRNA by priming with
 dt-tailed vector. The cDNA vector was circularized with
 have about 60nt. The cDNA vector was circularized with E.
 coli DNA ligase after digestion of EcoRI which site is
 also included in vector. An RNA strand converted to a DNA
 strand by Okayama-Berg method. The obtained cDNA vectors
 were used for transformation of competent cells E. coli
 Top10F' by electroporation method. The cDNA libraries
 constructed by this method are full-length enriched cDNA
 library."
 BASE COUNT 92 a 199 c 147 g 91 t
 ORIGIN

Query Match 64.6%; Score 514.8; DB 14; Length 529;
 Best Local Similarity 99.4%; Pred. No. 6e-102; 2; Indels 1; Gaps 1;
 Matches 527; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 31 CTCCTCACTCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 90
 Db 1 CTCCTCACTCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
 QY 91 TGGACCTCCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 150
 Db 61 TGGACCTCCGAGCA-GGGAGCTGTGGGGAGGAGGAGGAGGAGGAGGAGG 119
 QY 151 ACCTCCCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 210
 Db 120 ACCTCCCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 179
 QY 211 AGAAAAACCTATGACCTACCCGCTACCTGAGGAGGAGGAGGAGGAGGAGG 270
 Db 180 AGAAAAACCTATGACCTACCCGCTACCTGAGGAGGAGGAGGAGGAGGAGG 239
 QY 271 ATCTGAACCTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 330
 Db 240 ATCTGAACCTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 299
 QY 331 CAGAGACTTGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 390
 Db 300 CAGAGACTTGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 359
 QY 391 TGGCGCTGACCCAGAACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 450
 Db 360 TGGCGCTGACCCAGAACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 419
 QY 451 ACCGTGAGGCTGACCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 510
 Db 420 ACCGTGAGGCTGACCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 479
 QY 511 AGGGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 560
 Db 480 AGGGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 529

RESULT 13

BM846748
 LOCUS BM846748 569 bp mRNA linear EST 06-MAR-2002

DEFINITION	K-EST0125824 S13KMS5 Homo sapiens cDNA clone S13KMS5-40-G02 5', mRNA sequence.
ACCESSION	BM846748
VERSION	BM846748.1 GI:19203147
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 569)
AUTHORS	Kim,M.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
TITLE	21C Frontier Korean EST Project 2001
JOURNAL	Unpublished (2002)
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 40, row: G column: 02 High quality sequence stop: 569.
FEATURES	<p>Location/Qualifiers</p> <p>1..569</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="S13KMS5-40-G02"</p> <p>/clone_lib="S13KMS5"</p> <p>/tissue_type="myeloma"</p> <p>/cell_line="KMS-5"</p> <p>/lab_host="Top10P."</p> <p>/note="vector: pCNS; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10P, by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."</p>
BASE COUNT	135 a 192 c 151 g 91 t
ORIGIN	
	Query Match 57.6%; Score 458.8; DB 14; Length 569; Best Local Similarity 99.6%; Pred. No. 9.4e-90; Matches 460; Conservative 0; Mismatches 2; Indels 0; Gaps 0
Oy	1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCCGGGAGAGGAG 60
Db	108 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCCGGGAGAGGAG 167
Oy	61 CCGCACCCGCGCCCGAGCCCGAGCCCCATGGACCTCGAGCAGGGGACTCGTGGGGGA 120
Db	168 CCGCACCCGCGCGCCCGAGCCCCATGGACCTCGAGCAGGGGACTCGTGGGGGA 227
Oy	121 TGTTAGCGTGCTGTGACGGTGCTTGGCACCTCCTCGAGTCCAGTCCCAATCGCA 180
Db	228 TGTTAGCGTGCTGTGACGGTGCTTGGCACCTCCTCGAGTCCAGTCTCTCAATCGCA 287
Oy	181 CAGGGGACCCAGGCGCTTGCGCCCTCCATCCAGAAAACTATGACCTCACCCGCTACTCG 240
Db	288 CAGGGGACCCAGGCGCTTGCGCCCTCCATCCAGAAAACTATGACCTCACCCGCTACTCG 347
Oy	241 AGCACCAACTCCGAGCTTGCGTGGGACCTTACTGAACTACCTGGGCGCCCTTTCAACG 300

```

Db 61 TTAGTGGGACCTACCTGAACTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCT 120
Qy 318 CCCGCTCTGGGGGAGAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGGCGAGC 377
Db 121 CCTCGACTGGGGGAGAGAACTCTGCCAGGGCCACGGTCACTTGGAAAGTGTGGCGAAGC 180
Qy 378 CTCATGACAACTGGCGGTGACCCAGAGAACTACGAGGCTTACGCCACCTTCTGTGTAC 437
Db 181 CTCATGACAGGCTGGCGGTGACCCAGAACTATGAGGCGTACAGTCACTCTGTGTAC 240
Qy 438 TTGCGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGTGGCGGCGGAGGCTGGCCACTTC 497
Db 241 TTGCGTGGCTCAACCGCTCAGGCTGCCACTGCTGAGTGGCGGCGGAGGCTGGCCACTTC 300
Qy 498 TGACACAGCTCTCAGGCGCTGTGGGAGCATTTGGGCGCTCATGGCAGCTCTGGGCTAC 557
Db 301 TGTACACAGCTCTCAGGCGCTGTGGGAGCATTTGGGCGCTCATGGGCGCTGGGCTAC 360
Qy 558 CCACTGCCCCAGCGCTGCTGGGAGCATTTGGGCGCTCATGGCAGCTCTGGGCTAC 617
Db 361 CCACTGCCCCAGCGCTCTGGGAGCATTTGGGCGCTCATGGGCGCTGGGCTAC 420
Qy 618 GACTTCTCCAGAGATGGAGCTTCTGGCTGTGAGGAGCTGCGAGACCTGGCT 673
Db 421 GACTTCTCCAGAGATGGAGCTTCTGGCTGTGAGGAGCTGCGAGACCTGGCT 476

```

RESULT 15

```

BM764238
LOCUS
DEFINITION
K-EST0045718 S13KMS5 Homo sapiens cDNA clone S13KMS5-22-G02 5',
420 bp mRNA linear EST 04-MAR-2002
BM764238
ACCESSION
BM764238
VERSION
BM764238.1 GI:19093853
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 420)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 22 row: G column: 02
High quality sequence stop: 420.
Location/Qualifiers
1..420
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S13KMS5-22-G02"
/clone_lib="S13KMS5"
/tissue_type="myeloma"
/cell_line="RMS-5"
/lab_host="Top10F"
/note="vector; pcNS; Site 1: EcoRI; Site 2: NotI; The poly
phosphatase (BAP) and then decapped with bacterial alkaline
pyrophosphatase (TAP). The decapped intact mRNA was
ligated with DNA-RNA linker including EcoR I site by
treatment of T4 RNA ligase and the first strand cDNA was
synthesized from oligo dT-selected mRNA by priming with
dT-tailed vector. The dT-tailed vector was adjusted to
have about 60nt. The cDNA vector was circularized with E.

```

FEATURES source

coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

```

BASE COUNT      76 a 158 c 114 g 71 t      1 others
ORIGIN

Query Match
Best Local Similarity 50.9%; Score 405.4; DB 14; Length 420;
Matches 417; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 31 CTCCACTCCGAGCAGCCCTCCGGAGAGAGCGCCACCCGCGCGGCCGCCGCCGCCGCCA 90
Db 1 CTCCACTCCGAGCAGCCCTCCGGAGAGAGCGCCACCCGCGCGGCCGCCGCCGCCGCCA 60
Qy 91 TGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCCTGTGCACGGTGTCTGGC 150
Db 61 TGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCCTGTGCACGGTGTCTGGC 120
Qy 151 ACCTCCCTGCACTGCGAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATCC 210
Db 121 ACCTCCCTGCACTGCGAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATCC 180
Qy 211 AGAARACCTATGACCTCAACCCGCTACCTGGAGCACCAATCTCCGAGCTTGGCTGGGACCT 270
Db 181 AGAARACCTATGACCTCAACCCGCTACCTGGAGCACCAATCTCCGAGCTTGGCTGGGACCT 240
Qy 271 ATCTGAACCTACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGGG 330
Db 241 ATCTGAACCTACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGGG 300
Qy 331 CAGAGACTCTGCCCGAGGGCCAATGTTGACTTGGAGGTGGGNAAGCCTCAATGACAAAC 390
Db 301 CAGAGACTCTGCCCGAGGGCCAATGTTGACTTGGAGGTGGGNAAGCCTCAATGACAAAC 360
Qy 391 TCGCGCTGACCCAGAACTACGAGGCTACAG-CCACCTTCTGTGTACTTGGTGGCCTC 449
Db 361 TCGCGCTGACCCAGAACTACGAGGCTTACAGCCCACTTCTGTGTACTTGGTGGCCTC 420

```

Search completed: February 1, 2003, 12:58:32
Job time : 1626.85 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 202.966 Seconds
(without alignments)
8843.084 Million cell updates/sec

Title: US-09-931-704-1
Perfect score: 797
Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgctccccccc 797

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	797	100.0	797	19 AAV47510	Human neurotrophic
2	797	100.0	797	19 AAV22652	cDNA encoding huma
3	797	100.0	797	21 AAA39481	Human NNT-1 cDNA.
4	797	100.0	797	24 ABK11647	Human cDNA encodin
5	793.4	99.5	881	22 AAH74484	Nucleotide sequenc
6	793.4	99.5	1790	21 AAA88546	Human interleukin-
7	759.2	95.3	768	22 AAH99772	Human protein enco
8	749.4	94.0	1710	20 AAX16161	Human cardiotoxop
9	710	89.1	729	22 AAD04201	Human cardiotoxop

10	694.8	87.2	1008	22	AAKS1548	Human polynucleoti
11	669.4	84.0	819	19	AAV47512	Mouse neurotrophic
12	669.4	84.0	819	19	AAV22654	cDNA encoding muri
13	669.4	84.0	819	21	AAA39483	Murine NNT-1 cDNA.
14	669.4	84.0	819	24	ABK11649	Mouse cDNA encodin
15	668.4	83.9	968	22	ABA09140	Human cardiotoxop
16	668.4	83.9	968	22	AAKS2532	Human polynucleoti
17	558.4	70.1	648	21	AAA88547	Mouse interleukin-
18	523.4	65.7	5087	21	AAA39482	Human NNT-1 DNA.
19	523.4	65.7	5087	24	ABK11648	Human novel neurot
20	523.4	65.7	5088	19	AAV47511	Human neurotrophic
21	523.4	65.7	5088	19	AAV22653	Human genomic DNA
22	493.4	61.9	495	22	ABA17773	Human foetal liver
23	493.4	61.9	495	22	ABA37855	Probe #16321 for g
24	493.4	61.9	495	22	AAK20142	Human brain expres
25	493.4	61.9	495	22	AAK46202	Human bone marrow
26	493.4	61.9	495	22	AAI25564	Probe #15497 for g
27	493.4	61.9	495	22	AAI52108	Probe #20794 used
28	493.4	61.9	495	24	ABS20513	Human genome-deriv
29	369	46.3	492	22	ABA59256	Human foetal liver
30	369	46.3	492	22	ABA27996	Human brain expres
31	369	46.3	492	22	AAK07469	Human brain expres
32	369	46.3	492	22	AAK33253	Human bone marrow
33	369	46.3	492	22	AAI16384	Probe #8317 for ge
34	369	46.3	492	22	AAI139047	Probe #7733 used t
35	369	46.3	492	24	ABS08085	Human genome-deriv
36	269.4	33.8	283	22	AAH23145	Osteoarthritis tis
37	80.2	10.1	396	20	AAX16162	Human cardiotoxop
38	73.6	9.2	1260	24	ABQ16902	Oligonucleotide fo
39	73.6	9.2	1260	24	ABQ16903	Oligonucleotide fo
40	58.6	7.4	65	24	ABNS6698	Mouse spliced tran
41	49	6.1	627	20	ABQ60797	Human DNAX interle
42	46.8	5.9	1260	24	ABQ16904	Oligonucleotide fo
43	46.8	5.9	1260	24	ABQ16905	Oligonucleotide fo
44	46.8	5.9	2320	21	AAH21879	Human breast and o
45	46.2	5.8	627	20	AAH60796	Human DNAX interle

ALIGNMENTS

RESULT 1
ID AAV47510
AAV47510 standard; cDNA; 797 BP.

AC AAV47510;

DT 09-NOV-1998 (first entry)

DE Human neurotrophic factor NNT-1 cDNA.

KW NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW peripheral neuropathy; dystrophy; neural retina degeneration;
KW common variable immunodeficiency; CVID; selective IgA deficiency;
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
KW therapy; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 90...767

FT sig_peptide /tag= a

FT mat_peptide 90...170

FT mat_peptide 171...764

FT mat_peptide /tag= c

FN WO9833922-A1.

PD 06-AUG-1998.

XX Disclosure; Fig 1; 41pp; English.

CC The present sequence encodes a human neurotrophic factor, designated

CC NNT-1, which is capable of stimulating growth of motor or sympathetic

CC neurons. The NNT-1 protein is useful in the treatment of neurodegenerative

CC diseases characterized by the degeneration and death of particular

CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,

CC stroke and various degenerative disorders affecting vision.

XX

XX Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

XX

Query Match 100.0%; Score 797; DB 19; Length 797;

Best Local Similarity 100.0%; Pred. No. 1.1e-176;

Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCGAGCCGCGGCTCGCCCTCCCACTCCGAGCGGACTCGTGGGGA 60

DB 1 ATTAAGCTTCGCGAGCCGCGGCTCGCCCTCCCACTCCGAGCGGACTCGTGGGGA 60

QY 61 CCGCACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120

DB 61 CCGCACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120

QY 121 TGTAGCGTGTCTGTGACGCGTGTCTGCGACCTTCCCTCCGAGCGGACTCGTGGGGA 180

DB 121 TGTAGCGTGTCTGTGACGCGTGTCTGCGACCTTCCCTCCGAGCGGACTCGTGGGGA 180

QY 181 CAGGGGACCGAGGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

DB 181 CAGGGGACCGAGGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

QY 241 AGCACCAACTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300

DB 241 AGCACCAACTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300

QY 301 AGCCAGACTTCAACCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360

DB 301 AGCCAGACTTCAACCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360

QY 361 TGGAGTGTGGCAAGCGCTCAATGACAACTGCGGCGGCGGCGGCGGCGGCGGCGG 420

DB 361 TGGAGTGTGGCAAGCGCTCAATGACAACTGCGGCGGCGGCGGCGGCGGCGGCGG 420

QY 421 GCCACCTTCTGTACTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

DB 421 GCCACCTTCTGTACTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

QY 481 GCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540

DB 481 GCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540

QY 541 TGGGAGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

DB 541 TGGGAGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

QY 601 TTGGCCCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660

DB 601 TTGGCCCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660

QY 661 TGCAGACTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720

DB 661 TGCAGACTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720

QY 721 CTCAGCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780

DB 721 CTCAGCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780

QY 781 CCTCTTGGCTCCCCCCC 797

DB 781 CCTCTTGGCTCCCCCCC 797

RESULT 3

AAA39481
ID AAA39481 standard; cDNA; 797 BP.

AC AAA39481;

DT 24-AUG-2000 (first entry)

XX Human NNT-1 cDNA.

XX NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;
XX anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
XX nervous system degeneration; Alzheimer's disease; Parkinson's disease;
XX amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
XX Huntington's disease; peripheral neuropathy; neural retina degeneration;
XX retinopathy; immune disorder; hematopoietic disorder; ss.
XX Homo sapiens.

PH Key Location/Qualifiers
FT CDS 90..767
FT /*tag= a
FT /product= "NNT-1"

XX US6054294-A.

XX 25-APR-2000.

XX 12-DEC-1997; 97US-0988819.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M;

XX WPI; 2000-338492/29.

XX P-PSDB; AAY87813.

XX New nucleic acids encoding neurotrophic factors useful for stimulating
XX growth of motor or sympathetic neurons for treating neuron cell damage

XX Claim 1a; Fig 1; 42pp; English.

XX This invention describes a novel nucleic acid molecule (I) encoding a
XX novel neurotrophic factor (NNT-1) (ii) which has neurotrophic,
XX neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
XX ophthalmological activity. (I) is useful for producing NNT-1
XX polypeptides which are useful for treating patients in whom various
XX cells of the central, autonomic, or peripheral nervous system have
XX degenerated and/or have been damaged by congenital disease, trauma,
XX mechanical damage, surgery, stroke, ischemia, infection, metabolic
XX disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
XX proteins are used to treat diseases like Alzheimer's, Parkinson's,
XX amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
XX disease, peripheral neuropathy induced by diabetes or other metabolic
XX disorders, and/or dystrophies or degeneration of the neural retina such
XX as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
XX night blindness, progressive cone-rod degeneration, immune disorders and
XX hematopoietic disorders. (I) is effective in treating neurological
XX conditions and promotes neuron regeneration. Neural functions are
XX effectively restored in patients suffering from various neurological
XX disorders. This sequence encodes the human NNT-1 protein described in the
XX method of the invention.

XX Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

XX Query Match 100.0%; Score 797; DB 21; Length 797;

XX Best Local Similarity 100.0%; Pred. No. 1.1e-176;

XX Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sun Feb 2 08:31:37 2003

inflammatory bowel disease; transplant rejection; reproductive disorder; graft versus host disease; infertility; miscarriage; preterm labour.

Homo sapiens.

Key Location/Qualifiers
90..767
/*tag= a
/product= "NNT1"

WO200215977-A2.

28-FEB-2002.

17-AUG-2001; 2001WO-US25906.

18-AUG-2000; 2000US-226436P.

16-AUG-2001; 2001US-0931704.

(AMGE-) AMGEN INC.

Senaldi G;

WPI; 2002-280867/32.

P-PSDB; AAU78176.

Treating Immunoglobulin E-related disease, modulating IgE levels in a patient, preventing IgE-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient

Claim 2; Fig 1; 63pp; English.

The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor.

The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular stenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence encodes human NNT1.

Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;

Query Match 100.0%; Score 797; DB 24; Length 797;
Best Local Similarity 100.0%; Pred. No. 1.1e-176;
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCGAGGAGGAG 60
Db 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCGAGGAGGAG 60
QY 61 CGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 CGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 TGTAGCGTGCCTGTGACGCGTCTGGCACTCCCTGAGTCCAGCTCTCAATCGCA 180

QY 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCGAGGAGGAG 60
Db 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCGAGGAGGAG 60
QY 61 CGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 CGCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 TGTAGCGTGCCTGTGACGCGTCTGGCACTCCCTGAGTCCAGCTCTCAATCGCA 180
Db 121 TGTAGCGTGCCTGTGACGCGTCTGGCACTCCCTGAGTCCAGCTCTCAATCGCA 180
QY 181 CAGGGACCCAGGCGCTGGCGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240
Db 181 CAGGGACCCAGGCGCTGGCGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240
QY 241 AGCACAACTCCGAGCTTGGCTGGGAGCTTATCTGAATACCTGGGCGCCCTTCAACG 300
Db 241 AGCACAACTCCGAGCTTGGCTGGGAGCTTATCTGAATACCTGGGCGCCCTTCAACG 300
QY 301 AGCAGACTTCACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGCGCACTGTGACT 360
Db 301 AGCAGACTTCACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGCGCACTGTGACT 360
QY 361 TGGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCAAGTACGAGGCTTACA 420
Db 361 TGGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCAAGTACGAGGCTTACA 420
QY 421 GCCACCTTCTGTACTTGTGCTGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCC 480
Db 421 GCCACCTTCTGTACTTGTGCTGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCC 480
QY 481 GCAGCTGGCCCACTTCTGCACACGCTCCAGGGCTGTGGGAGCAATGCGGGCTCA 540
Db 481 GCAGCTGGCCCACTTCTGCACACGCTCCAGGGCTGTGGGAGCAATGCGGGCTCA 540
QY 541 TGGCAGCTTGGGCTACCACTGCGCCAGCCCTGCTGGGAGCAATGCGGGCTCA 600
Db 541 TGGCAGCTTGGGCTACCACTGCGCCAGCCCTGCTGGGAGCAATGCGGGCTCA 600
QY 601 CTGGCCCTGCCACAGTACTTCTCCAGAGTGAAGCTTCTGGCTGCTGAAGGAGC 660
Db 601 CTGGCCCTGCCACAGTACTTCTCCAGAGTGAAGCTTCTGGCTGCTGAAGGAGC 660
QY 661 TGCAGACTTGGCTGTGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGAGC 720
Db 661 TGCAGACTTGGCTGTGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGAGC 720
QY 721 CTCAGCAGCTGAGTCACTGAGTGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Db 721 CTCAGCAGCTGAGTCACTGAGTGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
QY 781 CCTCTTCGCTCCCGCCC 797
Db 781 CCTCTTCGCTCCCGCCC 797

RESULT 4

ABK11647

ID ABK11647 standard; cDNA; 797 BP.

XX ABK11647;

AC ABK11647;

05-JUN-2002 (first entry)

Human cDNA encoding novel neurotrophic factor NNT1.

Human; ss; gene; NNT1; neurotrophic factor; IgE-related disease; Type I allergic disease; allergic rhinitis; eczema; dermatitis; pollinosis; asthma; immune disease; cancer; arteriosclerosis; vascular stenosis; rheumatoid arthritis; psoriatic arthritis; inflammatory arthritis; osteoarthritis; inflammation joint disease; autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;

Db 121 TTTTGGGCTGCTGTGACGGTGTCTGGCAGCTCTCCCTGAGTGGCAGCTCTCAATCGCA 180
 Qy 181 CAGGGGACCCAGGGGCTGGCCCTTCCATCCAGAAACCTATGACCTCACCCTACCTGG 240
 Db 181 CAGGGGACCCAGGGGCTGGCCCTTCCATCCAGAAACCTATGACCTCACCCTACCTGG 240
 Qy 241 AGCACCAGCTCGAGCTGGCTGGGAGCTATCTGAATACCTACCTGGGCCCCCTTTCAAGC 300
 Db 241 AGCACCAGCTCGAGCTGGCTGGGAGCTATCTGAATACCTACCTGGGCCCCCTTTCAAGC 300
 Qy 301 AGCACCAGCTCGAGCTGGCTGGGAGCTATCTGAATACCTACCTGGGCCCCCTTTCAAGC 360
 Db 301 AGCACCAGCTCGAGCTGGCTGGGAGCTATCTGAATACCTACCTGGGCCCCCTTTCAAGC 360
 Qy 361 TGGAGGTGTGGGAGCTCGCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCTTACA 420
 Db 361 TGGAGGTGTGGGAGCTCGCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCTTACA 420
 Qy 421 GCCACCTTCTGTGTACTTGTGGCTCAACCGTCAAGCTGCGCTGCGCTGCGCTGCGCC 480
 Db 421 GCCACCTTCTGTGTACTTGTGGCTCAACCGTCAAGCTGCGCTGCGCTGCGCTGCGCC 480
 Qy 481 GCAGCTGGGCCACTTCTGACCAAGCTTCCAGGCGCTGCTGGGAGCATTTGGGGCGTCA 540
 Db 481 GCAGCTGGGCCACTTCTGACCAAGCTTCCAGGCGCTGCTGGGAGCATTTGGGGCGTCA 540
 Qy 541 TGGCAGCTTGGGCTACCACTGCGCCAGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 600
 Db 541 TGGCAGCTTGGGCTACCACTGCGCCAGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 600
 Qy 601 CTGGCCCTGCGCCAGCTGCTTCTCCAGAGATGGAGCTTCTGGCTGCTGAGGAGC 660
 Db 601 CTGGCCCTGCGCCAGCTGCTTCTCCAGAGATGGAGCTTCTGGCTGCTGAGGAGC 660
 Qy 661 TGCAGAGCTGGCTGCTGCGCTGCGCCAGGAGCTTCAACCGGCTCAAGAAAGATGAGC 720
 Db 661 TGCAGAGCTGGCTGCTGCGCTGCGCCAGGAGCTTCAACCGGCTCAAGAAAGATGAGC 720
 Qy 721 CTCACAGCTGGCTGCTGCGCTGCGCCAGGAGCTTCAACCGGCTCAAGAAAGATGAGC 780
 Db 721 CTCACAGCTGGCTGCTGCGCTGCGCCAGGAGCTTCAACCGGCTCAAGAAAGATGAGC 780
 Qy 781 CTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 797
 Db 781 CTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 797

RESULT 5

ID AAH74484
 AC AAH74484 standard; DNA; 881 BP.
 AC AAH74484;

15-OCT-2001 (first entry)
 DE Nucleotide sequence of a human NNT-1 protein.

KW NNT-1; CLF-1; SCNTFRalpha; nervous system; neuron; nervous system;
 KW neuro-muscular function; tumour; immune system; haematopoietic system;
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;
 KW blastocyst implantation; thrombosis; retinal disease;
 KW retinal pigmentosis; ss.

OS Homo sapiens.

Key Location/Qualifiers
 CDS 174..851
 FT /*tag= a
 FT /product= "NNT-1"
 XX

PN WO200155172-A2.
 XX 02-AUG-2001.
 PD 26-JAN-2001; 2001WO-FR00253.
 PF 27-JAN-2000; 2000FR-0001035.
 XX 12-OCT-2000; 2000FR-0013089.
 PR (FABR) FABRE MEDICAMENT SA PIERRE.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;
 DR WPI; 2001-488773/53.
 XX P-PSDB; AAG63543.
 DR A complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha
 PT protein useful to treat neurodegenerative disease including Parkinson's
 PT and Huntington's, obesity and cancer
 XX Disclosure; Page 57-58; 67pp; French.
 XX The present sequence encodes a human NNT-1 protein. The specification
 CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or
 CC SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate
 CC activity of the SCNTFRalpha/gp130/LiFbeta receptor complex, or to
 CC induce phosphorylation of the tyrosine of gp130 and LiFbeta, or to
 CC particularly where cells expressing the receptor complex are in the
 CC central or peripheral nervous system, in neurons implicated in
 CC neuro-muscular function or in skeletal muscle. The complex or
 CC antibodies are also used to decrease the survival, growth or
 CC proliferation of tumour cells or to facilitate the proliferation and/or
 CC inhibit differentiation of cells stocks. The complex is also used to
 CC modulate activity of the gp130/LiFbeta receptor or cells expressing
 CC that receptor, particularly those cells implicated in the immune,
 CC haematopoietic, nervous or reproductive system, the liver or skeletal
 CC muscle. Molecules of the invention may be used to prevent or treat
 CC neurodegenerative diseases including amyotrophic lateral sclerosis,
 CC Parkinson's and Huntington's disease, to repair or regenerate nervous
 CC or muscular tissue or to maintain muscular mass in paralytic patients.
 CC They may also be used to treat cancer, obesity and associated diseases,
 CC and to improve fertility, particularly to avoid endometriosis and/or
 CC assist blastocyst implantation, thrombosis, or retinal disease,
 CC particular retinal pigmentosis.
 XX SQ Sequence 881 BP; 158 A; 318 C; 246 G; 159 T; 0 other;

Query Match 99.5%; Score 793.4; DB 22; Length 881;
 Best Local Similarity 99.9%; Pred. No. 7,8e-176;
 Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ATTTAAGCTTCGCGGAGCGCGGCTCGCTTCCACTCCGAGCTCCGGAGAGGAG 60
 Db 85 ATTTAAGCTTCGCGGAGCGCGGCTCGCTTCCACTCCGAGCTCCGGAGAGGAG 144
 Qy 61 CGCACCCCGCGCGCCAGCCAGCCATGAGCTCCGAGCAGGAGACTCGTGGGGA 120
 Db 145 CGCACCCCGCGCGCCAGCCAGCCATGAGCTCCGAGCAGGAGACTCGTGGGGA 204
 Qy 121 TGTTAGCTGCTGTGACAGGTGTCTGGCACCTCCCTGCGAGTGCAGTCTCAATCGCA 180
 Db 205 TGTTAGCTGCTGTGACAGGTGTCTGGCACCTCCCTGCGAGTGCAGTCTCAATCGCA 264
 Qy 181 CAGGGGACCCAGGCGCTGGCCCCCTCCATCCAGAAACCTATGACCTACCGCTACCTGG 240
 Db 265 CAGGGGACCCAGGCGCTGGCCCCCTCCATCCAGAAACCTATGACCTACCGCTACCTGG 324
 Qy 241 AGCACCAACTCCGAGCTTGGCTGGGAGCTATCTGAACCTACCTGGGCCCCCTTTCAAGC 300
 Db 325 AGCACCAACTCCGAGCTTGGCTGGGAGCTATCTGAACCTACCTGGGCCCCCTTTCAAGC 384
 Qy 301 AGCCAGAGCTTCAACCCCTCCCGGCTGGGGGAGAGACTCTGCCCCAGGGGCCACTTGTGACT 360

Db	385	AGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACT	444
Qy	361	TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAACCCAGAACTAGAGGCTTACA	420
Db	445	TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAACCCAGAACTAGAGGCTTACA	504
Qy	421	GCCACCTTCTGTGTTACTTGGCTGCTCAACCGTCAGGCTGCCACTGCTGAGCTGGGCC	480
Db	505	GCCACCTTCTGTGTTACTTGGCTGCTCAACCGTCAGGCTGCCACTGCTGAGCTGGGCC	564
Qy	481	GCAGCTGGCCCACTTCTGACCAAGCTCCAGGGCTGCTGGGAGCACTTGGGGCTTCA	540
Db	565	GCAGCTGGCCCACTTCTGACCAAGCTCCAGGGCTGCTGGGAGCACTTGGGGCTTCA	624
Qy	541	TGGAGCTTCTGGGCTACCCACTGCCCGAGCGCTGCTGGGACTGAACCCACTTGGACTC	600
Db	625	TGGAGCTTCTGGGCTACCCACTGCCCGAGCGCTGCTGGGACTGAACCCACTTGGACTC	684
Qy	601	CTGGCCCTGCCACAGTACTTCTCCAGAGATGGAGACTTCTGGGCTGCTGAAGGAGC	660
Db	685	CTGGCCCTGCCACAGTACTTCTCCAGAGATGGAGACTTCTGGGCTGCTGAAGGAGC	744
Qy	661	TGCAGACTGCTGTGGGCTCGGCCAGGACTTCAACCGGCTCAAGAAAGATGCGAC	720
Db	745	TGCAGACTGCTGTGGGCTCGGCCAGGACTTCAACCGGCTCAAGAAAGATGCGAC	804
Qy	721	CTCCAGAGCTGAGTACCTCGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT	780
Db	805	CTCCAGAGCTGAGTACCTCGACCTGGGGGCTCATGGCTTCTGACTTCTGACTTCT	864
Qy	781	CTCTCTGCTCCCCC 795	
Db	865	CTCTCTGCTCCCCC 879	
RESULT 6			
AA88546	AA88546 standard; DNA; 1790 BP.		
XX	AC	AA88546;	
XX	XX	22-JAN-2001 (first entry)	
DE	XX	Human interleukin-B60 (IL-B60) gene.	
XX	XX	Interleukin-B60; IL-B60; human; cytokine; chromosome 11;	
KW	XX	cytokine-like factor-1; haematopoietic; inflammation;	
KW	XX	antiinflammatory; autoimmune disease; therapy; ds.	
XX	OS	Homo sapiens.	
XX	Key	Location/Qualifiers	
FH	CDS	162..809	
FT	/*tag= a		
FT	sig_peptide	162..212	
FT	/*tag= b		
FT	mat_peptide	213..806	
FT	/*tag= c		
XX	XX	WO20053631-A1.	
PN	XX	14-SEP-2000.	
XX	XX	09-MAR-2000; 2000WO-US06182.	
PF	XX	11-MAR-1999; 99US-0267901.	
XX	XX	(SCHE) SCHERING CORP.	
XX	XX	Oppmann B, Timans JC, Kastelein RA, Bazan JF;	
XX	XX	WPI; 2000-587426/55.	

DR	P-PSDB; AAB19686.		
XX	Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,		
PT	polypeptides, and nucleic acids, useful in research, diagnosis and for		
PT	treating inflammatory and autoimmune disorders -		
XX	Claim 17; Page 15-16; 97pp; English.		
XX	The present sequence is that of DNA encoding human interleukin-B60		
CC	(IL-B60, see AAB19586), a novel, small soluble cytokine-like protein		
CC	of 198 amino acids that exhibits structural motifs characteristic		
CC	of a member of the long-chain cytokines, and which shows homology		
CC	to granulocyte colony stimulating factor and interleukin-6. IL-60B		
CC	may have either stimulatory or inhibitory effects on haematopoietic		
CC	cells, including e.g. lymphoid cells, such as T-cells, B-cells,		
CC	natural killer cells, macrophages, dendritic cells, haematopoietic		
CC	progenitors, etc. Methods are provided for modulating the		
CC	physiology or development of a cell or tissue culture cells by		
CC	contacting the cell with an agonist or antagonist of IL-B60 or an		
CC	agonist of antagonist of a complex of mature IL-B60 and its		
CC	partner, cytokine-like factor-1 (CLF-1, see AAB19588). The		
CC	IL-B60/CLF-1 cytokine serves as a key physiological factor in motor		
CC	neuron development and regeneration. IL-60B, its agonists and		
CC	antagonists may be used to treat inflammatory or autoimmune		
CC	disorders and also for drug screening. The IL60B gene maps to		
CC	human chromosome 11.		
XX	Sequence 1790 BP; 381 A; 560 C; 474 G; 375 T; 0 other;		
Qy	Query Match	99.5%; Score 793.4; DB 21; Length 1790;	
Db	Best Local Similarity	99.9%; Pred. No. 9.1e-176;	
Qy	Matches 794; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1	ATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCACTCCGCGAGCTCCGGGAGGAG 60	
Db	43	ATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCACTCCGCGAGCTCCGGGAGGAG 102	
Qy	61	CCGACACCG 120	
Db	103	CCGACACCG 162	
Qy	121	TGTTAGCGTGCCTGTGCACGGTGTCTGGCACCTCCCTTCGAGTGCAGCTCAATCGCA 180	
Db	163	TGTTAGCGTGCCTGTGCACGGTGTCTGGCACCTCCCTTCGAGTGCAGCTCAATCGCA 222	
Qy	181	CAGGGGACCCAGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240	
Db	223	CAGGGGACCCAGGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282	
Qy	241	AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACTACCTGGGCGCGCGCG 300	
Db	283	AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACTACCTGGGCGCGCGCG 342	
Qy	301	AGCCAGACTTCAACCCCTCCCGCGCTGGGGGAGAGACTTGCCCGAGGGCCACTGTTGACT 360	
Db	343	AGCCAGACTTCAACCCCTCCCGCGCTGGGGGAGAGACTTGCCCGAGGGCCACTGTTGACT 402	
Qy	361	TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAACCCAGAACTAGAGGCTTACA 420	
Db	403	TGGAGGTGGGGAAGCCCTCAATGACAAACTGGGGTGAACCCAGAACTAGAGGCTTACA 462	
Qy	421	GCCACCTTCTGTGTTACTTGGCTGGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGGGCC 480	
Db	463	GCCACCTTCTGTGTTACTTGGCTGGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGGGCC 522	
Qy	481	GCAGCTGGCCCACTTCTGACCAAGCTCCAGGGCTGCTGGGAGCACTTGGGGCGGTCA 540	
Db	523	GCAGCTGGCCCACTTCTGACCAAGCTCCAGGGCTGCTGGGAGCACTTGGGGCGGTCA 582	
Qy	541	TGGAGCTTGGGCTACCCACTGCCCGAGCGCTGCTGGGAGCTGAACCCACTTGGGACTC 600	
Db	583	TGGAGCTTGGGCTACCCACTGCCCGAGCGCTGCTGGGAGCTGAACCCACTTGGGACTC 642	

QY 601 CTGGCCCTGCCACAGTCTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGC 660
 Db 643 CTGGCCCTGCCACAGTCTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGC 702
 QY 661 TGCAGACCTGGCTGTGGCGCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720
 Db 703 TGCAGACCTGGCTGTGGCGCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 762
 QY 721 CTCAGACCTGAGTACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
 Db 763 CTCAGACCTGAGTACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 822
 QY 781 CCTCTTGGCTCCCC 795
 Db 823 CCTCTTGGCTCCCC 837

RESULT 7
 AAH99772
 ID AAH99772 standard; cDNA; 768 BP.
 AC AAH99772;
 DT 16-OCT-2001 (first entry)
 XX Human protein encoding cDNA sequence SEQ ID NO:607.
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnary; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antinaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder; ss.
 XX Homo sapiens.
 OS
 XX
 XX WO200153455-A2.
 XX
 XX 26-JUL-2001.
 XX
 XX 22-DEC-2000; 2000WO-US35017.
 XX
 XX 23-DEC-1999; 99US-0471275.
 XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-457603/49.
 XX P-PSDB; AAM25831.
 XX
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 XX
 XX Claim 1; Page 638; 1217pp; English.
 XX
 XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 XX AAM25963. The proteins can have activities based on the tissues and
 XX cells they are expressed in, such as: antiinflammatory; antirheumatic;
 XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
 XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;

CC antiulcer; osteopathic; dermatological; antiallergic; antiaschmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

XX SQ Sequence 768 BP; 137 A; 278 C; 217 G; 136 T; 0 other;

Query Match 95.3%; Score 759.2; DB 22; Length 768;
 Best Local Similarity 99.6%; Pred. No. 7.3e-168;
 Matches 761; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATTAAGCTTCGCCGAGCGCGCTCGCCCTCCACCTCCGCGACGCTCCGCGAGGAG 60
 Db 5 ATTAAGCTTCGCCGAGCGCGCTCGCTTCCACTCCGCGACGCTCCGCGAGGAG 64

QY 61 CCGCACCG 120
 Db 65 CCGCACCG 124

QY 121 TGTAGCGTGTCTGTGACGCGTGTCTGTGGACCTCCCTGCGAGTCCGAGCTCTCAATCGCA 180
 Db 125 TGTAGCGTGTCTGTGACGCGTGTCTGTGGACCTCCCTGCGAGTCCGAGCTCTCAATCGCA 184

QY 181 CAGGGACCCAGCGCGCTGGCGCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 240
 Db 185 CAGGGACCCAGCGCGCTGGCGCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTGG 244

QY 241 AGCACCAACTCCGCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGCGCGCGCGCGCG 300
 Db 245 AGCACCAACTCCGCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGCGCGCGCGCGCG 304

QY 301 AGCCAGACTTCAACCTCCCGCGCTGGGGCGAGAGCTCTGCCAGCGCGCGCGCGCGCTG 360
 Db 305 AGCCAGACTTCAACCTCCCGCGCTGGGGCGAGAGCTCTGCCAGCGCGCGCGCGCGCTG 364

QY 361 TGGAGGTGTGGCGAAGCGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGCGCTTACA 420
 Db 365 TGGAGGTGTGGCGAAGCGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGCGCTTACA 424

QY 421 GCCACCTTCTGTGTACTTGTGGCTCAACCGTCAGGCTGCCACTGTGAGCTGGCGCC 480
 Db 425 GCCACCTTCTGTGTACTTGTGGCTCAACCGTCAGGCTGCCACTGTGAGCTGGCGCC 484

QY 481 GCAGCTTGGCGCACTTCTGCAACAGCTTCCAGGGCTGTGGGCGAGCATTTGGCGGCTCA 540
 Db 485 GCAGCTTGGCGCACTTCTGCAACAGCTTCCAGGGCTGTGGGCGAGCATTTGGCGGCTCA 544

QY 541 TGGCAGCTCTGGCGTACCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
 Db 545 TGGCAGCTCTGGCGTACCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 604

QY 601 CTGGCGCTGCCACAGTGTCTTCTCCAGAGATGACGACTTCTGGCTGCTGAGGAGC 660
 Db 605 CTGGCGCTGCCACAGTGTCTTCTCCAGAGATGACGACTTCTGGCTGCTGAGGAGC 664

QY 661 TGCAGACCTGGCTGTGGCGCTCGGCGCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720
 Db 665 TGCAGACCTGGCTGTGGCGCTCGGCGCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 724

QY 721 CTCAGACCTGAGTACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 764
 Db 725 CTCAGACCTGAGTACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 768

FT /product= "Human mature cardiostrophin-like cytokine
 XX (CLC) protein"
 PN WO200127157-A1.
 XX 19-APR-2001.
 XX 06-OCT-2000; 2000WO-AU01216.
 XX 08-OCT-1999; 99AU-0003327.
 PR 12-MAY-2000; 2000AU-0007489.
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;
 PI Nakata Y, Hasegawa M;
 XX WPI; 2001-281978/29.
 DR P-PSDB; AAE00828.
 XX New biologically active complex comprising NR6 and
 PT cardiostrophin-like-cytokine, for facilitating proliferation,
 PT differentiation and/or survival of a cell -
 XX Claim 31; Page 112-114; 123pp; English.
 XX The present invention relates to a biologically active complex comprising
 CC a haemopoietin receptor, NR6 and cardiostrophin-like cytokine (CLC).
 CC The complex is useful in the manufacture of a medicament for the
 CC treatment and/or prophylaxis of a subject, as it is involved in
 CC facilitating proliferation, differentiation and/or survival of a cell.
 CC The complex or its components have neurotrophic activity. The present
 CC sequence is human cardiostrophin-like cytokine (CLC) cDNA.
 XX Sequence 729 BP; 132 A; 261 C; 196 G; 140 T; 0 other;
 SQ

Query Match 89.1%; Score 710; DB 22; Length 729;
 Best Local Similarity 100.0%; Pred. No. 2.2e-156;
 Matches 710; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 CCCCATGGACCTCCGAGCGGGGACTCGTGGGGGATGTTAGCGTGTGCGACGGTGTCT 145
 Db 7 CCCCATGGACCTCCGAGCGGGGACTCGTGGGGGATGTTAGCGTGTGCGACGGTGTCT 66
 QY 146 CTGGACCTCCCTGCGAGTCCAGCTCTCAATCGCAGGGGACCCAGGCGCTGGCCCCCTC 205
 Db 67 CTGGACCTCCCTGCGAGTCCAGCTCTCAATCGCAGGGGACCCAGGCGCTGGCCCCCTC 126
 QY 206 CATCCAGAAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCTTGGCTGG 265
 Db 127 CATCCAGAAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCTTGGCTGG 186
 QY 266 GACCTATCTGAATCTGAGTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCT 325
 Db 187 GACCTATCTGAATCTGAGTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCT 246
 QY 326 GGGGCGAGAGTCTGCCCCAGGGCCACTGTTCACCTTGGAGGTGCGGAGGCTCAATGA 385
 Db 247 GGGGCGAGAGTCTGCCCCAGGGCCACTGTTCACCTTGGAGGTGCGGAGGCTCAATGA 306
 QY 386 CAACCTGGGCTGACCCAGAACTAGAGCCCTACAGCCACCTTCTGTGTACTTGGCTGG 445
 Db 307 CAACCTGGGCTGACCCAGAACTAGAGCCCTTACAGCCACCTTCTGTGTACTTGGCTGG 366
 QY 446 CCTCAACCGTTCAGGCTGACCTGCTGAGCTGGCGCGAGCTGGGCCCACTTCTGACACGAG 505
 Db 367 CCTCAACCGTTCAGGCTGACCTGCTGAGCTGGCGCGAGCTGGGCCCACTTCTGACACGAG 426
 QY 506 CCTCAGGGCTGCTGGGAGCATTGCGGGGCTCATGCGAGCTCTGGGCTACCCACTGCC 565
 Db 427 CCTCAGGGCTGCTGGGAGCATTGCGGGGCTCATGCGAGCTCTGGGCTACCCACTGCC 486
 QY 566 CCAGCGCTGCTGGGAGTGAACCACTTGGACTCTCTGGGCCCTGCCCACTGACTTCTCT 625

Db 487 CCAGCGCTGCTGGAGTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTGAATCTCT 546
 QY 626 CCAGAAGATGACGACTTCTGGCTGCTGAAGAGCTGAGACCTGGCTGTGGCGCTCGGC 685
 Db 547 CCAGAAGATGACGACTTCTGGCTGCTGAAGAGCTGAGACCTGGCTGTGGCGCTCGGC 606
 QY 686 CAAGGACTTCAACCGGCTCAAGAGAGATGACGCTCCAGCAGCTGACCTGCA 745
 Db 607 CAAGGACTTCAACCGGCTCAAGAGAGATGACGCTCCAGCAGCTGACCTGCA 666
 QY 746 CCTGGGGCTCATGGCTTCTGACTTCTGACCTTCTGCTCTCTCTCTCTCTCTCTCT 795
 Db 667 CCTGGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCT 716

RESULT 10
 AAK51548
 ID AAK51548 standard; cDNA; 1008 BP.
 XX
 AC AAK51548;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 93.
 XX
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR P-PSDB; AAM78415.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 711-712; 622pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK51435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

XX sequence 1008 BP: 183 A; 374 C; 247 G; 204 T; 0 other;
SO

Query Match	87.2%;	Score 694.8;	DB 22;	Length 1008;
Best Local Similarity	97.0%;	Pred. No. 8.3e-153;		
Matches 708; Conservative	0;	Mismatches 22;	Indels 0;	Gaps 0;

66	CCGCGCGCGCCAGCCCGAGCCCATGACCTCCGAGCAGGGGACTCGTGGGGATGTTA	125
267	CCCGGGCGCGCTCATGTCTTCGTCCATCCTGCCACAGGGGACTCGTGGGGATGTTA	326
126	GGTGCCCTGTGCACGGTGCTTGGCACTTCCTCGAGTGCCAGCTCTCAATCGCACAGG	185
327	GGTGCCCTGTGCACGGTGCTTGGCACTTCCTCGAGTGCCAGCTCTCAATCGCACAGG	386
186	GACCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTGGAGCAC	245
387	GACCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCCACCGCTACCTGGAGCAC	446
246	CAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTGGGCCCTTTCAACAGACCA	305
447	CAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTGGGCCCTTTCAACAGACCA	506
306	GACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAG	365
507	GACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACTTTGGAG	566
366	GTGTGGGAGGCTCAATGACAAACTCGGGCTCACCCAGAACTACGAGGCTACAGCCAC	425
567	GTGTGGGAGGCTCAATGACAAACTCGGGCTGACCCAGAACTACGAGGCTACAGCCAC	626
426	CTTCTGTGTACTTGGTGGCCCTCAACCGTCAGGTGCGCACTGCTGAGCTGCCCGGAGC	485
627	CTTCTGTGTACTTGGTGGCCCTCAACCGTCAGGTGCGCACTGCTGAGCTGCCCGGAGC	686
486	CTGGCCCACTTTGCGCACAGGCTCCAGGCGCTGCTGGGAGCACTTGGGCGCTCATGGCA	545
687	CTGGCCCACTTTGCGCACAGGCTCCAGGCGCTGCTGGGAGCACTTGGGCGCTCATGGCA	746
546	GCTCTGGGCTACCCACTGCCAGCGCTGCTGGGACTGAACCCACTTGGACTCTTGGC	605
747	GCTCTGGGCTACCCACTGCCAGCGCTGCTGGGACTGAACCCACTTGGACTCTTGGC	806
606	CCTGCCCACTGTGACTTCTCTCCAGAAATGACGACTTCTGGCTGCTGAAGGAGCTGCAG	665
807	CCTGCCCACTGTGACTTCTCTCCAGAAATGACGACTTCTGGCTGCTGAAGGAGCTGCAG	866
666	ACCTGGCTGTGGCGCTGGGCCAAGGACTTCAACGGCTCAAGAGAGATGAGGCTTCCA	725
867	ACCTGGCTGTGGCGCTGGGCCAAGGACTTCAACGGCTCAAGAGAGATGAGGCTTCCA	926
726	GCAGTGCAGTCAACCTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCT	785
927	GCAGTGCAGTCAACCTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCT	986
786	TCGCTCCCC	795
987	TCGCTCCCC	996

RESULT 11
AAV47512
ID AAV47512 standard: cDNA: 819 BP:

AAV47512:

09-NOV-1998 (first entry)

XX
Mouse neurotrophic factor NNT-1 cDNA.

XX 22

NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;
 Alzheimer's disease, Parkinson's disease; Huntington's disease;
 amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
 peripheral neuropathy; dystrophy; neural retina degeneration;
 common variable immunodeficiency; CVID; selective IgA deficiency;
 hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
 therapy; ss.

Mus sp.

Location/Qualifiers

Key	95..772
CDS	

/*tag= a

sig_peptide	95%.175	1st30%	h

mat nentide
/ *Lay= 0
176.769

```
mac_per_tag
/*tag= c
```

WO9833922-A1.

06-AUG-1998.

02-FEB-1998: 98WO-US02363.

30-JAN-1998; 98US-0016534.

03-FEB-1997; 97US-0792019.

(AMGE-) AMGEN INC.

Chang M Elliot GS. Sarmiento U, Senaldi G;

WDT: 1998-437475/37.

P-PSDB; AAW29716.

Newly isolated nucleic acid encoding human or murine neurotrophic factor NNT-1 - useful for treatment of neurological and inflammatory or inflammation, also as vaccine adjuvant

Disclosure: Fig 4: 120pp: English:

This newly isolated mouse cDNA sequence codes for a novel neurotrophic factor, designated NNT-1 (see AAW29716), that is a growth factor for neurons and for B or T cells. Vectors and host cells are provided for use in the production of murine and human NNT-1 polypeptides. These are used to treat: (i) neurological or immunological diseases, specifically Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral sclerosis, dystrophy and Charcot-Marie-Tooth syndrome, peripheral neuropathy, degeneration of the neural retina, or conditions characterised by T or B cell defects, e.g. common variable immunodeficiency (CVID), selective IgA deficiency, hypogammaglobulinaemia and X-linked agammaglobulinaemia (claimed), but many others disclosed; and (ii) inflammation. NNT-1 is also able to boost immunoreactivity and antibody production following vaccination, and, since it inhibits tumour necrosis factor production, it may also be useful for treating sepsis. NNT-1 nucleic acid fragments are also used as hybridisation probes in diagnostic assays. In addition, cells that have been engineered to express NNT-1 can be implanted, or nucleic acids can be delivered in gene therapy vectors.

2 010 201 150 A. 200 C. 218 G. 157 T. 0 other:

Query Match 84.0%; Score 669.4; DB 19; Length 819;
Best Local Similarity 92.0%; Pred. NO. 6.8e-147;

1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGCAGCCTCCGGAGAGGAG 60
2 ATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGCAGCCTTCGGAGAGGAG 64
3 ATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGCAGCCTTCGGAGAGGAG 68
4 ATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGCAGCCTTCGGAGAGGAG 72

61 CCGCACCCGGCCGGCCAG - CCCCAGCCCCATGGACCTCCGAGCAGGGGACTCGTGGGG 119
65 CCGGCCCCGGGCGGCCGGCCCCCAGCCCCATGGACCTCCGAGCAGGGGACTCGTGGGG 124


```

QY 120 ATGTTAGCGTCTGTGCGAGGTGCTTGGACCTCTCCAGCTCTCCAGTGCAGCTCTCAATCGC 179
Db 125 ATGTTAGCGTCTGTGCGAGGTGCTTGGACCTCTCCAGTGCAGCTCTCAATCGC 184
QY 180 ACAGGGGACCCAGGGGCTGGCCCTCCATCCAGAAACCTATGACCTACCCGCTACCTG 239
Db 185 ACAGGAGATCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTACCCGCTACCTG 244
QY 240 GAGCACCACCTCGCAGCTTGGCTGGGACCTATCTGAACTACTCTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCACTCCGAGCTTAGCTGGGACCTACCTGAACTACTCTGGGCCCCCTTTCAAC 304
QY 300 GAGCCAGACTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCCCCACTGTGTGAC 359
Db 305 GAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGCCCCAGGTCAAC 364
QY 360 TTGGAGGTGTGGCGAAGCTCAATGACAACTCGCGGCTGACCCAGAACTACGAGGCTTAC 419
Db 365 TTGGAAGTGTGGCGAAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCTAC 424
QY 420 AGCCACCTCTGTGTACTTGGTGGCTCAACCTCAGGCTGCCACTGCTGAGCTGGC 479
Db 425 AGTCACCTCTGTGTACTTGGTGGCTCAACCTCAGGCTGCCACTGCTGAGCTGGC 484
QY 480 CGCAGCTTGGCCCACTCTGACCAAGCTCCAGGCTGCTGGGAGCAATTCGAGGTGTC 544
Db 485 CGTAGGCTGGCCCACTCTGACCAAGCTCCAGGCTGCTGGGAGCAATTCGAGGTGTC 549
QY 540 ATGGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGTC 599
Db 545 ATGGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGTC 604
QY 600 CTTGGCCCTTGGCCAGCTTCTCCAGAGATGGAAGACTTCTGGCTGCTGAAGGAG 659
Db 605 CTTGGCCCTTGGCCAGCTTCTCCAGAGATGGAAGACTTCTGGCTGCTGAAGGAG 664
QY 660 CTGAGACTTGGCTGGGCTGGGCTGGGCAAGGACTTCAACCGGCTCAAGAGAGATGCGAG 719
Db 665 CTGAGACTTGGCTGGGCTGGGCTGGGCAAGGACTTCAACCGGCTCAAGAGAGATGCGAG 724
QY 720 CTTCCAGAGCTGAGTACCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTT 778
Db 725 CTTCCAGAGCTTCACTGACCTGGAGGCAATGGTTTCTGACCTTGAACCT 783

RESULT 12
AAV22654
ID AAV22654 standard; cDNA; 819 BP.
XX
AC AAV22654;
XX
DT 13-JUL-1998 (first entry)
XX
DE cDNA encoding murine neurotrophic factor NNT-1.
XX
KW Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 95..771
FT sig_peptide /tag= a
FT 95..175 /tag= b
FT mat_peptide 176..769 /tag= c
XX
FT US5741772-A.
XX
PD 21-APR-1998.

```

```

XX 03-FEB-1997; 97US-0792019.
XX 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
XX Chang M;
XX WPI; 1998-260526/23.
XX P-PSDB; AAW56142.
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
XX useful for stimulating growth of motor and sympathetic neurons
XX Disclosure; Fig 4; 41pp; English.
XX The present sequence encodes a murine neurotrophic factor, designated
XX NNT-1, which is capable of stimulating growth of motor or sympathetic
XX neurons. The NNT-1 protein is useful in the treatment of neurological
XX diseases characterized by the degeneration and death of particular
XX classes of neurons. These diseases specifically include Parkinson's
XX disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
XX stroke and various degenerative disorders affecting vision.
XX Sequence 819 BP; 156 A; 288 C; 218 G; 157 T; 0 other;
XX
Query Match 84.0%; Score 669.4; DB 19; Length 819;
Best Local Similarity 92.0%; Pred. No. 6.8e-147;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY 1 ATTAAGCTTTCGGCGAGCGCGCTCGCCCTCCACTCCGCGAGCTTCGGGAGAGGAG 60
Db 5 ATTAAGCTTTCGGCGAGCGCGCTCGCCCTCCACTCCGCGAGCTTCGGGAGAGGAG 64
QY 61 CCGCAACCGCGCCCGCCAG--CCCGAGCCCATGAGCTCCGAGCAGGAGCTCGTGGGG 119
Db 65 CCGCGCCCGCGCGCGCGCGCCCGAGCCCATGAGCTCCGAGCAGGAGCTCGTGGGG 124
QY 120 ATCTTACGCTGCTGTGTCAGGCTGTGGCACTCCCTGAGTGGCAGCTCTCAATCGC 179
Db 125 ATCTTACGCTGCTGTGTCAGGCTGTGGCACTCCCTGAGTGGCAGCTCTCAATCGC 184
QY 180 ACAGGGGACCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCTACCCGCTACCTG 239
Db 185 ACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCTACCCGCTACCTG 244
QY 240 GAGCACCACCTCGCAGCTTGGCTGGGACCTATCTGAACTACTCTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCACTCCGAGCTTAGCTGGGACCTACCTGAACTACTCTGGGCCCCCTTTCAAC 304
QY 300 GAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCGCCAGGCGCACTGTGAC 359
Db 305 GAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGCGCACTGGTCAAC 364
QY 360 TTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTACGAGGCTTAC 419
Db 365 TTGGAAGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGAACTATGAGGCTTAC 424
QY 420 AGCCACCTCTGTGTACTTGGTGGCTCAACCTGAGGTGCGCACTGCTGAGCTGGCG 479
Db 425 AGTCACCTCTGTGTACTTGGTGGCTCAACCTGAGGTGCGCACTGCTGAGCTGGCG 484
QY 480 CGCAGCTTGGCCCACTTCTGACCAAGCTCCAGGCTGCTGGGAGCAATTCGAGGTGTC 539
Db 485 CGTAGGCTGGCCCACTTCTGACCAAGCTCCAGGCTGCTGGGAGCAATTCGAGGTGTC 544
QY 540 ATGGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGTC 599
Db 545 ATGGAGCTTGGGCTACCCACTGCCAGCGCTGCTGGGAGCAATTCGAGGTGTC 604
QY 600 CTTGGCCCTTGGCCAGCTTCTCCAGAGATGGAAGACTTCTGGCTGCTGAAGGAG 659

```


KW	Mouse; ss; Gene; NNT1; neurotrophic factor; IGF-related disease;
KW	Type I allergic disease; allergic rhinitis; eczema; dermatitis;
KW	pollinosis; asthma; immune disease; cancer; arteriosclerosis;
KW	vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
KW	inflammatory arthritis; osteoarthritis; inflammatory joint disease;
KW	autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
KW	inflammatory bowel disease; transplant rejection; reproductive disorder;
KW	graft versus host disease; infertility; miscarriage; preterm labour.
XX	
OS	Mus SP.

OS Mus sp.

Key	Location/Qualifiers
FH	
Key	
Location/Qualifiers	

FT	CDS	95..7
FT		

```

FF      /*tag=, a
FF      /

```

```

FT
XX
/product="NNT1"

```

XX
PN
WO200215977-A2.

XX
PD 28-FEB-2002.

20-1EE-2002,
XX
PF 17-AUG-2001; 2001WO-US25906.

XX
XX

0965750-OWTVOO (2002) 2002-05-08

PR 18-AUG-2000; 2000US-226436P.
PR 16-AUG-2001; 2001US-0931704.
.....

XX
PA (AMGE-) AMGEN INC.XX
PI Senaldi G;XX
 General G;

DR WPI; 2002-280867/
DR P-PSDB; AAU78177.

XX
PT Treating Immunoglobulin E-related disease, modulating IgE levels in a
PT patient, preventing IgE-related disease and treating allergic diseases,
PT involves administering NNT-1 inhibitor to a patient -
PS
PS Claim 2; Fig 4; 63pp; English.

PS Claim 2; Fig 4; 63pp; English.

The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor. The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular stenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence encodes mouse NNT1.

Sequence 819 BP; 156 A; 288 C; 218 G; 157 T; 0 other;
SQ

Query Match

Query Match	Score	DB	Length
Best Local Similarity	84.0%	669.4	24
Best Local Similarity	92.0%	pred	pred

best local similarity 92.0%; Pred. No. 6.8e-147;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 1 ATTAAGCTTCGCGAGCGGGCTCGCCCTCCACTCGCCAGCCTCGGAGAGGAG 60

Dp 5 ATTAAGCTTCGCGAGCGGGCTCGCCCTCCACTCGCCAGCCTCGGAGAGGAG 60

Db
5 ATTAAGCTTCGCGGAGCGGGCTCGCCCTCCCACTCGCGAGCCTCTGGAGAGGAG 64

	Query Match	83.9%	Score 668.4	DB 22	Length 968
Best Local Similarity	57.0%		Pred. NO. 1.2e-146		
Matches 681	Conservative 0		Mismatches 21		Indels 0
					Gaps 0
66	CCCGCCGCCGCCAGCCAGCCCGCTCGAGCAGGGACTCGTGGGGATGTTA	125			
267	CCCGCCGCCGGCTCATGCTTCTCGTCCATCTCGGCCACAGGGGACTCGTGGGGATGTTA	326			

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 09:12:22 ; Search time 34.2438 Seconds
(without alignments)
10456.327 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 797

Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgctcccccc 797

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	797	100.0	797	10	US-09-931-704-1
2	669.4	84.0	819	10	US-09-931-704-4
3	523.4	65.7	5087	10	US-09-931-704-3
4	493.4	61.9	495	10	US-09-864-761-23175
5	369	46.3	492	10	US-09-864-761-6462
6	269.4	33.8	283	10	US-09-765-231A-75
7	44.8	5.6	68750	9	US-10-014-717-1
8	38.4	4.9	460	10	US-09-864-761-2889
9	38.4	4.8	1324	10	US-09-925-301-363
10	38	4.8	11336	9	US-09-764-868-1440
11	37.8	4.7	3129	10	US-09-925-697-1
12	37.4	4.7	3088	10	US-09-954-456-45
13	37.4	4.7	3088	10	US-09-954-456-1621
14	37.4	4.7	3088	10	US-09-969-347-234
15	37.2	4.7	3984	9	US-09-954-531-151
16	37.2	4.7	4559	10	US-09-919-172-61
17	37.2	4.7	4559	12	US-10-044-090-555
18	36.8	4.6	4462	9	US-09-974-298-28
19	36.8	4.6	4469	10	US-09-796-008-1

c 20	36.8	4.6	4695	9	US-09-945-901-57	Sequence 57, Appl
c 21	36.8	4.6	4695	9	US-10-007-747-57	Sequence 57, Appl
c 22	36.8	4.6	4695	9	US-10-038-937-57	Sequence 57, Appl
c 23	36.4	4.6	1434	10	US-09-822-830A-53	Sequence 53, Appl
c 24	36.4	4.6	3466	9	US-09-880-464-1	Sequence 1, Appl
c 25	36.4	4.6	3466	10	US-09-338-405-1	Sequence 1, Appl
c 26	36.2	4.5	444	10	US-09-864-761-1526	Sequence 1526, Ap
c 27	35.6	4.5	292	10	US-09-050-516-16	Sequence 16, Appl
c 28	35.6	4.5	1841	10	US-09-050-516-3	Sequence 3, Appl
c 29	35.6	4.5	2894	10	US-09-050-516-27	Sequence 27, Appl
c 30	35.6	4.5	3098	9	US-09-981-353-96	Sequence 96, Appl
c 31	35.2	4.4	394	10	US-09-983-965-2488	Sequence 2488, Ap
c 32	35.2	4.4	474	10	US-09-864-761-5659	Sequence 5659, Ap
c 33	35.2	4.4	1554	10	US-09-799-777-103	Sequence 103, App
c 34	35.2	4.4	2943	10	US-09-764-878-306	Sequence 306, App
c 35	35.2	4.4	2947	10	US-09-764-878-304	Sequence 304, App
c 36	35.2	4.4	2953	10	US-09-764-878-302	Sequence 302, App
c 37	35	4.4	2327	10	US-09-954-456-1159	Sequence 1159, Ap
c 38	35	4.4	15500	10	US-09-764-869-2109	Sequence 2109, Ap
c 39	34.8	4.4	451	10	US-09-781-867-3	Sequence 3, Appl
c 40	34.8	4.4	6160	9	US-10-029-413A-11	Sequence 11, Appl
c 41	34.8	4.4	6160	9	US-10-029-413A-13	Sequence 13, Appl
c 42	34.8	4.4	6615	9	US-10-029-413A-9	Sequence 9, Appl
c 43	34.8	4.4	152331	9	US-10-035-407-16	Sequence 16, Appl
c 44	34.6	4.3	1785	10	US-09-864-761-18040	Sequence 18040, A
c 45	34.6	4.3	2428	10	US-09-822-849A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-09-931-704-1
; Sequence 1, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using
; TITLE OF INVENTION: Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (90)..(764)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (171)..()
; OTHER INFORMATION: Met at -27
; US-09-931-704-1

Query Match 100.0%; Score 797; DB 10; Length 797;
Best Local Similarity 100.0%; Pred. No. 4.7e-181;
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATTAAGCTTCGCGGAGCGGCTCGCTCCCTCCACTCCGCGGAGGAGGAG 60
Db	1	ATTAAGCTTCGCGGAGCGGCTCGCTCCCTCCACTCCGCGGAGGAGGAG 60
Qy	61	CGCACCACCGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGGAG 120
Db	61	CGCACCACCGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGGAG 120
Qy	121	TGTTAGCGTGGCTGTGCGAGCGGTGCTCTGGACCTCCCTCCAGTGCAGCTCTCAATCGCA 180
Db	121	TGTTAGCGTGGCTGTGCGAGCGGTGCTCTGGACCTCCCTCCAGTGCAGCTCTCAATCGCA 180

Db 121 TGTTAGCGTCTGTGACGGTGTCTGGCACCTCCCTGCGAGTGCAGCTCTCAATCGCA 180
Qy 181 CAGGGACCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACTCAACCGCTACCTGG 240
Db 181 CAGGGACCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACTCAACCGCTACCTGG 240
Qy 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
Db 241 AGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
Qy 301 AGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCATCTGTGACT 360
Db 301 AGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCATCTGTGACT 360
Qy 361 TGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACAGGCGCTACA 420
Db 361 TGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACAGGCGCTACA 420
Qy 421 GCCACCTTCTGTGTTACTTGGTGGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCGCC 480
Db 421 GCCACCTTCTGTGTTACTTGGTGGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGCGCC 480
Qy 481 GCAGCTTGGGCGACTTGTGACAGGCTTCCAGGCGCTGCTGGGAGCATTTGGGGCGTCA 540
Db 481 GCAGCTTGGGCGACTTGTGACAGGCTTCCAGGCGCTGCTGGGAGCATTTGGGGCGTCA 540
Qy 541 TGGCAGCTTGGGCTACCCACTGCCCCAGCGCTGCTGGGACTGAACCCACTTGGACTC 600
Db 541 TGGCAGCTTGGGCTACCCACTGCCCCAGCGCTGCTGGGACTGAACCCACTTGGACTC 600
Qy 601 CTGGCCCTGCCCCAGTGACTTCTCCAGAGATGGAGACTTCTGGCTGTGAAGGAGC 660
Db 601 CTGGCCCTGCCCCAGTGACTTCTCCAGAGATGGAGACTTCTGGCTGTGAAGGAGC 660
Qy 661 TGACAGCTTGGCTGTGGCGCTGGCGCAAGGACTTCAACCGGCTCAAGAGAGATGAGC 720
Db 661 TGACAGCTTGGCTGTGGCGCTGGCGCAAGGACTTCAACCGGCTCAAGAGAGATGAGC 720
Qy 721 CTCAGCAGCTGACGTCACTGACCTGCGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Db 721 CTCAGCAGCTGACGTCACTGACCTGCGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Qy 781 CTTCTTGGCTCCCCC 797
Db 781 CTTCTTGGCTCCCCC 797

RESULT 2

US-09-931-704-4

; Sequence 4, Application US/09931704

; Patent No. US20020041873A1

; GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio

; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using N

; FILE REFERENCE: A-695

; CURRENT APPLICATION NUMBER: US/09/931,704

; PRIOR FILING DATE: 2001-08-16

; PRIOR APPLICATION NUMBER: US 60/226,436

; PRIOR FILING DATE: 2000-08-18

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 819

; TYPE: DNA

; ORGANISM: Murine

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (95)..(769)

; OTHER INFORMATION:

; NAME/KEY: mat_peptide

; LOCATION: (176)..()

; OTHER INFORMATION:

; NAME/KEY: mat_peptide
; LOCATION: (176)..(769)
; OTHER INFORMATION:
; NAME/KEY: sig_peptide
; LOCATION: (95)..(175)
; OTHER INFORMATION:
US-09-931-704-4

Query Match 84.0%; Score 669.4; DB 10; Length 819;

Best Local Similarity 92.0%; Pred. No. 1.1e-150;

Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 1 ATTAAGCTTCCCGGAGCGCGGCTCGCCCTCCCACTCCGCCAGCTCCGGGAGAGGAG 60
Db 5 ATTAAGCTTCCCGGAGCGCGGCTCGCCCTCCCACTCCGCCAGCTCCGGGAGAGGAG 64
Qy 61 CGGACCCCGGCGGCCAG-CCCCAGCCCCATGAGACTCCGAGCAGGAGACTCGTGGGG 119
Db 65 CGGCGCGGCGGCCGCCGCCCATGGACCTCCGAGCAGGAGACTCGTGGGG 124
Qy 120 ATGTTAGCTGCTGTGACGGTGTCTGGCACCTCCCTGCGAGTGCAGCTCTCAATCGC 179
Db 125 ATGTTAGCTGCTGTGACGGTGTCTGGCACCTCCCTGCGAGTGCAGCTCTTAATCGC 184
Qy 180 ACAGGGGACCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTCAACCGCTACCTG 239
Db 185 ACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACCTCAACCGCTACCTG 244
Qy 240 GAGCACCACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCAACTCCGAGCTTGGCTGGGACCTACTGGAATCTACTGGGCCCCCTTTCAAC 304
Qy 300 GAGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGAC 359
Db 305 GAGCCTGACTTCAATCTCTCGACTGGGGGAGAAAATCTGCCAGGGCCACGCTCAAC 364
Qy 360 TTGGAGGTGTGGGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCGTAC 419
Db 365 TTGGAAGTGTGGGAGGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGCGTAC 424
Qy 420 AGCAGCTTCTGTGTTACTTGGTGGGCTCAACCGTCAAGCTGCGCTGAGCTGCGC 479
Db 425 AGTCACTCTCTGTGTACTTGGCTGGGCTCAACCGCTCAGGCTGCCACAGCTGAACCTCGA 484
Qy 480 CGCAGCTTGGCGGCTTCTGACAGGCTCCAGGGGCTCTGGGAGGACTTGGGGGCTC 539
Db 485 CGTAGCTTGGGCGGCTTCTGTACAGGCTCCAGGGGCTCTGGGAGGACTTGGAGGTTC 544
Qy 540 ATGGCAGCTTGGGCTTACCCACTGCCCCAGCGGCTGCTGGGACTGAACCACTTGGACT 599
Db 545 ATGGCGAGCTTGGCTTACCCACTGCCCCAGGCTCTGCCAGGACTGAGCCAGCTTGGGC 604
Qy 600 CTGGCCCTTGGCCAGTGAATCTCTCAGAGATGAGAGACTTCTGGCTGCTGAAGGAG 659
Db 605 CTGGCCCTTGGCCAGTGAATCTCTCAGAGATGAGAGACTTCTGGCTGCTGAAGGAG 664
Qy 660 CTGACAGCTGCTGTGGGCTCGGCAAGGACTTCAACCGGCTCAAGAGAGATGAGCAG 719
Db 665 CTGACAGCTGCTGTGGGCTTACGCAAGGACTTCAACCGGCTTAAAGAGAGATGAGCAG 724
Qy 720 CTTCCAGAGCTGAGTCACTCCCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTT 778
Db 725 CTTCCAGAGCTTCACTCACTCCCTGACCTTGGAGGCGACATGGTTTCTGACCTCTGACCCT 783

RESULT 3

US-09-931-704-3

; Sequence 3, Application US/09931704

; Patent No. US20020041873A1

; GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio

; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using

; TITLE OF INVENTION: Inhibitors


```
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 5087
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (137)..(138)
; OTHER INFORMATION: product = "INTERVENING UNSEQUENCED REGION OF >1KB"
US-09-931-704-3

Query Match      65.7%; Score 523.4; DB 10; Length 5087;
Best Local Similarity 99.8%; Pred. No. 1e-115;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 271 ATCTGAACCTACCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCCTGGGG 330
Db 363 AGCTGAACCTACCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCCTGGGG 3422
Qy 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 390
Db 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482
Qy 391 TCGCGCTGACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTACTTGTGGCCTCA 450
Db 3483 TCGCGCTGACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTACTTGTGGCCTCA 3542
Qy 451 ACCGTCAGGCTGCCACTGTGAGCTGCGCGCAGCTGCGCCACTTCTGCACAGCCTCC 510
Db 3543 ACCGTCAGGCTGCCACTGTGAGCTGCGCGCAGCTGCGCCACTTCTGCACAGCCTCC 3602
Qy 511 AGGGCTCTGTGGGAGCATTGGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC 570
Db 3603 AGGGCTCTGTGGGAGCATTGGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
Qy 571 CGCTCCTGGGACTGAACCCACTTGGACTCTGGGCTGCGCCACAGTGAATTCCTCCAGA 630
Db 3663 CGCTCCTGGGACTGAACCCACTTGGACTCTGGGCTGCGCCACAGTGAATTCCTCCAGA 3722
Qy 631 AGATGAGCAGACTTCTGGCTGTGAGGAGCTCAGACCTGGCTGTGGCTGGCCCAAGG 690
Db 3723 AGATGAGCAGACTTCTGGCTGTGAGGAGCTCAGACCTGGCTGTGGCTGGCCCAAGG 3782
Qy 691 ACTTCAACCGGCTCAAGAAGATGACAGCTCCAGCAGCTGCAGTCAACCTGCACCTGG 750
Db 3783 ACTTCAACCGGCTCAAGAAGATGACAGCTCCAGCAGCTGCAGTCAACCTGCACCTGG 3842
Qy 751 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3887

RESULT 4
US-09-864-761-23175/c
; Sequence 23175, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
```

```
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23175
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
; OTHER INFORMATION: NT HIT: g11439486, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: A1752561.1, EVALUE 0.00e+00
US-09-864-761-23175
```

Query Match 61.9%; Score 493.4; DB 10; Length 495;
Best Local Similarity 99.8%; Pred. No. 8e-109;
Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 273 CTGAACCTACCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCCTGGGGCA 332
Db 495 CTGAACCTACCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCCTGGGGCA 436
Qy 333 GAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAACTG 392
Db 435 AAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAACTG 376
Qy 393 CGGCTGACCCAGAACTACAGAGCCTACAGCCACTTCTGTGTACTTGTGGTGGCCTCAAC 452
Db 375 CGGCTGACCCAGAACTACAGAGCCTACAGCCACTTCTGTGTACTTGTGGTGGCCTCAAC 316
Qy 453 CGTCAGGCTGCCACTGTGCTGAGCTGCGCGCAGCCTGGCCACTTCTTCTGCACCAGCCTCCAG 512
```


[illegible]

RESULT 5
US-09-864-761-6462/c
; Sequence 6462, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

```

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6462
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4.5
;
; US-09-864-761-6462
;
; Query Match 46.3%; Score 369; DB 10; Length 492;
; Best Local Similarity 100.0%; Pred. No. 3.3e-79;
; Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 6
US-09-765-231A-75
; Sequence 75, Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
; APPLICANT: Searle/Monsanto
; APPLICANT: Phippard, Deborah
; APPLICANT: Vasanthakamur, Geetha
; APPLICANT: Dotson, Stanton
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides
; TITLE OF INVENTION: vectors, and cells
; FILE REFERENCE: SO-3221 PR
; CURRENT APPLICATION NUMBER: US/09765,231A
; CURRENT FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 75
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-765-231A-75

```

Query Match	33.8%;	Score 269.4;	DB 10;	Length 283;
Best Local Similarity	99.3%;	Pred. No. 1.5e-55;		
Matches 281;	Conservative 0;	Mismatches 1;	Indels 1;	Gaps 0;

QY	37	CTCCGCGACGCTCCGGGAGAGAGCGCAGCCGCGCGGCCGACCCAGCCCGCCAGCCCATGGACC	96
DB	1	CTCCGCGACGCTCCGGGAGAGAGCGCAGCCGCGCGGCCGACCCAGCCCGCCAGCCCATGGACC	60
QY	97	TCCGAGCAGGGGACT-CGTGGGGGATGTTAGCGTGCCTGTGCACGGTGTCTTGGCACCTC	155
DB	61	TCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCCTGTGCACGGTGTCTTGGCACCTC	120
QY	156	CCTGAGTGCAGCTCTCAATCGCACAGGGACCCAGGGCTGGGCCCTCCATCCAGAAA	215
DB	121	CCTGAGTGCAGCTCTCAATCGCACAGGGACCCAGGGCTGGGCCCTCCATCCAGAAA	180
QY	216	ACCTATGACCTCACCCGCTACTCTGGAGCACCAACTCCGACGTTGGCTGGGACCTATCTG	275
DB	181	ACCTATGACCTCACCCGCTACTCTGGAGCACCAACTCCGACGTTGGCTGGGACCTATCTG	240
QY	276	AACTACCTGGGCCCCCCTTTCAACGAGCGACACTTCAACCCCTC	318
DB	241	AACTACCTGGGCCCCCCTTTCAACGAGCGACACTTCAACCCCTC	283

RESULT 7

US-10-014-717-1

; Sequence 1, Application US/10014717

; Publication No. US20020192778A1

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/10/014,717

; CURRENT FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US/09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 1

; LENGTH: 68750

; TYPE: DNA

; ORGANISM: Sorangium cellulosum

US-10-014-717-1

Query Match	5.6%;	Score 44.8;	DB 9;	Length 68750;
Best Local Similarity	50.0%;	Pred. No. 0.16;		
Matches 112;	Conservative 0;	Mismatches 112;	Indels 0;	Gaps 0;

QY	456	CAGGCTGCCACTGTGAGCTGCGCGGAGCTGGGCCCACTTCTGCACAGGCTCCAGGCG	515
DB	57992	CAGGTAGCGAGTCGGGAGCGGGGAAAGCTGCGCGCGCTTCCGATCCAGCCGCGCGC	58051
QY	516	CTGCTGGGCGACATTGCGGGCGTTCATGGCAGTCTGGGCTACCCACTGCCCGACGCGGTG	575
DB	58052	GTGCTGCGCGGATCGGGCGGTTCAGACCCGCGGAGTGCACCTCGCGCCCTGGCG	58111
QY	576	CCTGGGACTGAACCCACTTGGAGCTCTGGGCCCTGCCCACTGATCTTCTCCAGAAATG	635
DB	58112	GCCCGGCTTCATGCGCGCGTCCCGCTTCGCGCTATCTATGTTGGCTCGCCGAGATGGG	58171
QY	636	GACGACTCTTGGCTGCTGAAGAGCTGCAGACCTGGCTGTGGCG	679
DB	58172	CTTCAATACGCGCGCGTTCGGGGGCTCGCCGAGCTGTGGCG	58215

RESULT 8

Qy 5 AGCTTCGCGAGCGCGGCTCGCCCTCCACTTCGCGCAGCTCTCCGGGAGAGAGCGCGCA 65
380 AGCTCTCCCCCAACTGCGCATGGGGTCTCTGCGCAGCCCAAGCTCTCTGGAGAGACACCGT 321
Db
Qy 66 CCGGCGCGGCCACGCCACCCCACTGGACCTCCGACAGGGACTCGTGGGGG 119
320 CCTGTCTCCACGGCACCCAGTCCCATCGACCGGCTAAGGCTCAGGAGTGCGGG 267
Db

```

RESULT 9
US-09-925-301-363
; Sequence 363, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 363
; LENGTH: 1324
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (385)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-925-301-363

```

```

RESULT 10
US-09-764-868-1440/c
; Sequence 1440, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1440
; LENGTH: 11336
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4205)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-1440

Query Match      4.8%; Score 38; DB 9; Length 11336;
Best Local Similarity 67.9%; Pred. No. 4.4;
Matches 53; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 412 AGGCTACAGCCACTTCTCTGTGTACTTGGTGGCTTCAACCGTCAAGGTCGCCACTGCTG 471
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1466 AGGCTCGGACACTCTGCAGTGGCTCCAGCTGGCCCCACCCACAGCCAGCCACTGCTG 1407

QY 472 AGCTGCGCGCAGCCTGG 489
      ||||| ||||| |||||
DB 1406 TCCTGCCCATAACTCGG 1389

RESULT 11
US-09-925-697-1/c
; Sequence 1, Application US/09925697
; Patent No. US20020119511A1
; GENERAL INFORMATION:
; APPLICANT: May, Earl
; APPLICANT: Van Horn, Stephanie
; APPLICANT: Warren, Patrick V.
; APPLICANT: Warren, Richard L.
; TITLE OF INVENTION: dnaE
; FILE REFERENCE: GM10237
; CURRENT APPLICATION NUMBER: US/09/925,697
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US/09/387,695
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3129
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-925-697-1

Query Match      4.7%; Score 37.8; DB 10; Length 3129;
Best Local Similarity 47.3%; Pred. No. 3.7;
Matches 114; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 479 CGCAGCTGGCGCCACTTCTGCACCAGCCTCCAGGGCTGCTGGCGAGCATTTGCGGGCGT 538
DB 2056 CTGCAAACTTCTCATAAACATCAAAACCTGCTGTGATTTTCCGAGATAGGCTGCTT 1997

QY 539 CATGGCAGCTCTGGGCTACCCACTGCCCGCCGCTGCTGGGACTGAACCCACTTGGAC 598
DB 1996 CTATGGAGCCTTGATAAAGGAGCCCTCATCTCATGATAGCAGAGCATCTTTTTCAC 1937

QY 599 TCCTGGCCCTGCCCACTGACTTCTCTCCAGAGATGAGCAGCTTCTGGCTGCTGAAGGA 658
DB 1936 CCATAGCTCGAGCAAAATATCGGCCCTTCCCAAGACTAAATCCAGCAAAATCGCTGAGCAA 1877

QY 659 GCTGAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAAGATGCA 718
DB 1876 CTTGCATAACCTGCTCTCTGATAGAGCATAAATGCCATAAGTTGGAGCCAAATATCTCTCA 1817

QY 719 G 719
DB 1816 G 1816

RESULT 12
US-09-954-456-45
; Sequence 45, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul

```

```

RESULT 11
US-09-925-697-1/c
; Sequence 1, Application US/09925697
; Patent No. US20020119511a1
; GENERAL INFORMATION:
; APPLICANT: May, Earl
; APPLICANT: Van Horn, Stephanie
; APPLICANT: Warren, Patrick V.
; APPLICANT: Warren, Richard L.
; TITLE OF INVENTION: dnaE
; FILE REFERENCE: GM10237
; CURRENT APPLICATION NUMBER: US/09/925,697
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US/09/387,695
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3129
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-925-697-1

```

RESULT 12
US-09-954-456-45
; Sequence 45, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul

RESULT 13
US-09-954-456-1621
; Sequence 1621, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25

	Query Match	4.7%	Score 37.4;	DB 10;	Length 3088;
	Best Local Similarity	52.8%;	Pred. No. 4.5;		
	Matches 103;	Conservative 0;	Mismatches 91;	Indels 1;	Gaps 1;
QY	135	TGCACGGTCTTGGCACCTCTCCAGTGTCCAGCTTCAATCGCACAGGGACCCAGGG	194		
Db	407	TGCACGACACCTTGGGCCCCCATCTTCAGAATCACTCCCTGCCGGTGTGCACAGGCCACGG	466		
QY	195	CTGTGGCCCTTCCATCCAGAAAACCTATGACCTCACCCGGTACTCTGGAGCACCAACTCCGC	254		

Db 467 CCAGCAGGATCC-CCTGGCAGAACTACCACTGAACGACTGGATGGAGGAGGAATACCGC 525
QY 255 AGCTTGGCTGGGACCTATCTGAATCTGCTGGCCCCCTTTCAACGAGCCAGACTTCAAC 314
Db 526 CACTTCCCGGGGAGTACGTCCTTACCGGCTACCCCTGCTCTGCTGAGACCTTCTACCAAC 585
QY 315 CCTCCCGGCTGGGG 329
Db 586 CACCTCCGCCAGGAG 600

RESULT 15

US-09-954-531-151/C
; Sequence 151, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 151
; LENGTH: 3984
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-151

Query Match 4.7%; Score 37.2; DB 9; Length 3984;
Best Local Similarity 52.6%; Pred. No. 5.4;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 560 ACTGCCCCAGCCGCTGGCTGGGACTGACCCACTTGGACTCCTGGCCCTGCCACAGTGA 619
Db 1081 ACTGGCCTCTGCGCATGCTCTCCGGCAGCCACTCGGCTCTGGCTATGCTCTCTGGGG 1022
QY 620 CTTCTCCAGAGATGGACGACTTCTGGCTGTGAAGGAGCTGCAGACCTGGCTGTGGCG 679
Db 1021 GACCACCTGCATGAGCTTCTGACACAGGACGGTCTGTGGAGAGATGCCCTGGAGGAGCCA 962
QY 680 CTGGCCAAGACTTCAACCGGCTCAAGAGAAG 713
Db 961 CTTGGGCTGTGTTTTCCACCAGACCCCGAGAAG 928

Search completed: February 2, 2003, 03:55:32
Job time : 102.244 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 25.6828 Seconds
(without alignments)
9516.918 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 797

Sequence: 1 attaaagcttcgcccggagcc.....tctctcttctgctccccccc 797

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	797	100.0	797	1	US-08-792-019B-1
2	797	100.0	797	3	US-08-988-819-1
3	797	100.0	797	3	US-09-016-534-1
4	749.4	94.0	1710	3	US-09-106-182-1
5	669.4	84.0	819	1	US-08-792-019B-4
6	669.4	84.0	819	3	US-08-988-819-4
7	669.4	84.0	819	3	US-09-016-534-4
8	523.4	65.7	5087	1	US-08-792-019B-3
9	523.4	65.7	5087	3	US-08-988-819-3
10	523.4	65.7	5087	3	US-09-016-534-3
11	80.2	10.1	396	3	US-09-106-182-7
12	44.8	5.6	68750	3	US-09-335-409-1
13	44.8	5.6	68750	4	US-09-568-102-1
14	44.8	5.6	68750	4	US-09-567-969-1
15	44.8	5.6	68750	4	US-09-568-480-1
16	44.8	5.6	68750	4	US-09-568-486-1
17	44.8	5.6	68750	4	US-09-568-472-1
18	44.8	5.6	68750	4	US-09-567-899-1
19	38.4	4.8	71989	4	US-09-443-501A-2
20	38	4.8	223	1	US-08-383-761-1
21	38	4.8	223	1	US-08-824-277-1
22	37.8	4.7	3129	4	US-09-387-695-1
23	37.8	4.7	43272	1	US-08-614-770A-1
24	36.8	4.6	1308	4	US-08-937-067-18
25	36.8	4.6	4695	2	US-08-231-193A-57
26	36.8	4.6	4695	2	US-08-486-273A-57
27	36.8	4.6	4695	3	US-08-940-086A-57

c	28	36.8	4.6	4695	4	US-08-940-035A-57	Sequence 57, Appl
c	29	36.8	4.6	4695	4	US-08-935-105A-57	Sequence 57, Appl
c	30	36.8	4.6	4695	4	US-09-648-797-57	Sequence 57, Appl
c	31	36.4	4.6	645	1	US-08-312-870-4	Sequence 4, Appl
	32	36.4	4.6	1338	1	US-08-307-444A-8	Sequence 8, Appl
	33	36.4	4.6	1338	1	US-08-587-389-8	Sequence 8, Appl
	34	36.4	4.6	1368	1	US-08-307-444A-7	Sequence 7, Appl
	35	36.4	4.6	1368	1	US-08-587-389-7	Sequence 7, Appl
	36	36.4	4.6	1425	1	US-08-307-444A-6	Sequence 6, Appl
	37	36.4	4.6	1425	1	US-08-307-444A-23	Sequence 23, Appl
	38	36.4	4.6	1425	1	US-08-587-389-6	Sequence 6, Appl
	39	36.4	4.6	1425	1	US-08-587-389-23	Sequence 23, Appl
	40	36.4	4.6	1428	1	US-08-014-723-3	Sequence 3, Appl
	41	36.4	4.6	1428	1	US-08-014-723-4	Sequence 4, Appl
	42	36.4	4.6	1428	1	US-08-110-011A-3	Sequence 3, Appl
	43	36.4	4.6	1428	1	US-08-110-011A-4	Sequence 4, Appl
	44	36.4	4.6	1491	1	US-08-312-870-2	Sequence 2, Appl
	45	36.4	4.6	1680	1	US-08-014-723-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-792-019B-1
; Sequence 1, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUTROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-08-792-019B-1

Query Match 100.0%; Score 797; DB 1; Length 797;

Best Local Similarity 100.0%; Pred. No. 3.2e-161; Mismatches 0; Indels 0; Gaps 0;
Matches 797; Conservative 0;

```
QY 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGCGAGCCCTCCGGGAGGAG 60
|
|
|
Db 1 ATTAAGCTTCGCGGAGCGCGCTCGCCCTCCACTCCGCGAGCCCTCCGGGAGGAG 60
|
|
|
QY 61 CCGACCCCGCGCGCCAGCCCCAGCCCATGGACCTCCGAGCGAGGACTCGTGGGGGA 120
|
|
|
Db 61 CCGACCCCGCGCGCCAGCCCCAGCCCATGGACCTCCGAGCGAGGACTCGTGGGGGA 120
|
|
|
QY 121 TGTAGCGTGTGTCACGGTGTCTGGCACTTCCGCACTCCCTCGAGTGCAGCTCTCAATCGCA 180
|
|
|
Db 121 TGTAGCGTGTGTCACGGTGTCTGGCACTTCCGCACTCCCTCGAGTGCAGCTCTCAATCGCA 180
|
|
|
QY 181 CAGGGAGCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTGG 240
|
|
|
Db 181 CAGGGAGCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTGG 240
|
|
|
QY 241 AGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
|
|
|
Db 241 AGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
|
|
|
QY 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCCACTGTTGACT 360
|
|
|
Db 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCCACTGTTGACT 360
|
|
|
QY 361 TGGAGGTGTGCGGAAGCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420
|
|
|
Db 361 TGGAGGTGTGCGGAAGCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420
|
|
|
QY 421 GCACCTTCTGTGTTACTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 480
|
|
|
Db 421 GCACCTTCTGTGTTACTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 480
|
|
|
QY 481 GCAGCCTGCGCCACTTCTGCACCCAGGCTCCAGGGCTGCTGGGCGAGCATTTGCGGCGTCA 540
|
|
|
Db 481 GCAGCCTGCGCCACTTCTGCACCCAGGCTCCAGGGCTGCTGGGCGAGCATTTGCGGCGTCA 540
|
|
|
QY 541 TGGCAGCTCTGGGCTACCACTGCGCCAGCGCTGCGCTGGGACTGAACCCACTTGGACTC 600
|
|
|
Db 541 TGGCAGCTCTGGGCTACCACTGCGCCAGCGCTGCGCTGGGACTGAACCCACTTGGACTC 600
|
|
|
QY 601 CTGGCCCTGCCAGAGTACTTCTCCAGAGATGGAGCACTTCTGGCTGTGAAGGAGC 660
|
|
|
Db 601 CTGGCCCTGCCAGAGTACTTCTCCAGAGATGGAGCACTTCTGGCTGTGAAGGAGC 660
|
|
|
QY 661 TGCAGACTGCTGTGGCTGCGGCTCGGCAAGGACTTCAACCGGCTCAAGAAGAGATGCAGC 720
|
|
|
Db 661 TGCAGACTGCTGTGGCTGCGGCTCGGCAAGGACTTCAACCGGCTCAAGAAGAGATGCAGC 720
|
|
|
QY 721 CTCAGAGCTGCACTCACTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
|
|
|
Db 721 CTCAGAGCTGCACTCACTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
|
|
|
QY 781 CCTCTTCTGCTCCCCCCC 797
|
|
|
Db 781 CCTCTTCTGCTCCCCCCC 797
|
|
|
```

RESULT 2

```
US-08-988-819-1
; Sequence 1, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
```

```
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
;
US-08-988-819-1
```

Query Match 100.0%; Score 797; DB 3; Length 797;

Best Local Similarity 100.0%; Pred. No. 3.2e-181; Indels 0; Gaps 0;

Matches 797; Conservative 0; Mismatches 0;

```
QY 1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCACTCCGCGAGCCCTCCGGGAGGAG 60
Db 1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCACTCCGCGAGCCCTCCGGGAGGAG 60
QY 61 CCGACCCCGCGCGCCAGCCCCAGCCCATGGACCTCCGAGCGAGGACTCGTGGGGGA 120
Db 61 CCGACCCCGCGCGCCAGCCCCAGCCCATGGACCTCCGAGCGAGGACTCGTGGGGGA 120
QY 121 TGTAGCGTGTGTCACGGTGTCTGGCACTTCCGCACTCCCTCGAGTGCAGCTCTCAATCGCA 180
Db 121 TGTAGCGTGTGTCACGGTGTCTGGCACTTCCGCACTCCCTCGAGTGCAGCTCTCAATCGCA 180
QY 181 CAGGGAGCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTGG 240
Db 181 CAGGGAGCCAGGCGCTCGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTGG 240
QY 241 AGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
Db 241 AGCAACAACTCCGAGCTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 300
QY 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCCACTGTTGACT 360
Db 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGAGAGACTCTGCCAGGGGCCACTGTTGACT 360
QY 361 TGGAGGTGTGCGGAAGCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420
Db 361 TGGAGGTGTGCGGAAGCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCTTACA 420
QY 421 GCACCTTCTGTGTTACTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 480
Db 421 GCACCTTCTGTGTTACTTGGCTGGGACCTATCTGAATCTACTGGGCCCCCTTTCAACG 480
QY 481 GCAGCCTGCGCCACTTCTGCACCCAGGCTCCAGGGCTGCTGGGCGAGCATTTGCGGCGTCA 540
Db 481 GCAGCCTGCGCCACTTCTGCACCCAGGCTCCAGGGCTGCTGGGCGAGCATTTGCGGCGTCA 540
|
|
|
```


Qy 541' TGGCAGCTCTGGGCTACCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600
Db 541 TGGCAGCTCTGGGCTACCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600
Qy 601 CTGGCCCTGCCCCAGCTACTTCTCCAGAGATGGAGCTTCTGGCTGCTGAAGGAGC 660
Db 601 CTGGCCCTGCCCCAGCTACTTCTCCAGAGATGGAGCTTCTGGCTGCTGAAGGAGC 660
Qy 661 TGCAGACTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720
Db 661 TGCAGACTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720
Qy 721 CTCAGAGCTGCTAGTCACTGACCTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Db 721 CTCAGAGCTGCTAGTCACTGACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Qy 781 CCTCTTGGCTCCCCCCC 797
Db 781 CCTCTTGGCTCCCCCCC 797

RESULT 3

US-09-016-534-1
; Sequence 1, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:

CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-09-016-534-1

Query Match 100.0%; Score 797; DB 3; Length 797;
Best Local Similarity 100.0%; Pred. No. 3.2e-181;
Matches 797; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATTTAAGCTTCCGCGAGCCGCGGCTCGCCCTCCCACTCCGCGACCTCCGCGAGAGAG 60
Db 1 ATTTAAGCTTCCGCGAGCCGCGGCTCGCCCTCCCACTCCGCGACCTCCGCGAGAGAG 60
Qy 61 CCGCACC CGGCGCGCCAGCCGCCAGCCCATGAGACCTCCGAGCAGGGGACTCGTGGGGGA 120
Db 61 CCGCACC CGGCGCGCCAGCCGCCAGCCCATGAGACCTCCGAGCAGGGGACTCGTGGGGGA 120
Qy 121 TGTAGCGTCTGTGACAGGTGTCTTGGCACCTCCCTGCACTGCGAGCTCTCAATCGCA 180
Db 121 TGTAGCGTCTGTGACAGGTGTCTTGGCACCTCCCTGCACTGCGAGCTCTCAATCGCA 180
Qy 181 CAGGGACCCAGGGGCTGGGCCCTCCATCCAGAAAACTATGACCTACCCGCTACCTGG 240
Db 181 CAGGGACCCAGGGGCTGGGCCCTCCATCCAGAAAACTATGACCTACCCGCTACCTGG 240
Qy 241 AGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCCTTTCAACG 300
Db 241 AGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCCTTTCAACG 300
Qy 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTGTACT 360
Db 301 AGCCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGACTCTGCCAGGGCCACTGTGTACT 360
Qy 361 TGGAGGTGTGGGAAGCCTCAATGACAACTGGGGCTGACCCAGAACTACGAGGCTTACA 420
Db 361 TGGAGGTGTGGGAAGCCTCAATGACAACTGGGGCTGACCCAGAACTACGAGGCTTACA 420
Qy 421 GCCACCTTCTGTGTACTTGTGGTGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGCC 480
Db 421 GCCACCTTCTGTGTACTTGTGGTGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGCC 480
Qy 481 GCAGCTGGCCCACTTCTGCACACCGCTCCAGGGCTCTCTGGGCGAGCATTTGGGGCGTCA 540
Db 481 GCAGCTGGCCCACTTCTGCACACCGCTCCAGGGCTCTCTGGGCGAGCATTTGGGGCGTCA 540
Qy 541 TGGCAGCTCTGGGCTACCCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600
Db 541 TGGCAGCTCTGGGCTACCCACTGCCCCAGCCGCTGCTGGGACTGAACCCACTTGGACTC 600
Qy 601 CTGGCCCTGCCCCAGAGTACTTCTCCAGAGATGGAGCTTCTGGCTGCTGAAGGAGC 660
Db 601 CTGGCCCTGCCCCAGAGTACTTCTCCAGAGATGGAGCTTCTGGCTGCTGAAGGAGC 660
Qy 661 TGCAGACTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720
Db 661 TGCAGACTGGCTGTGGGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATGCAGC 720
Qy 721 CTCAGAGCTGCTAGTCACTGACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Db 721 CTCAGAGCTGCTAGTCACTGACCTGCACTGGGGGCTCATGGCTTCTGACTTCTGACCTTCT 780
Qy 781 CCTCTTGGCTCCCCCCC 797
Db 781 CCTCTTGGCTCCCCCCC 797

RESULT 4
US-09-106-182-1
; Sequence 1, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yanggu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave

	CITY:	Rockville	
	STATE:	MD	
	COUNTRY:	US	
	ZIP:	20850	
	COMPUTER READABLE FORM:		
	MEDIUM TYPE:	Floppy disk	
	COMPUTER:	IBM PC compatible	
	OPERATING SYSTEM:	PC-DOS/MS-DOS	
	SOFTWARE:	PatentIn Release #1.0, Version #1.30	
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER:	US/09/106,182	
	FILING DATE:	Herewith	
	CLASSIFICATION:		
	PRIOR APPLICATION NUMBER:	US 60/051,053	
	APPLICATION NUMBER:	30-JUN-1997	
	FILING DATE:		
	ATTORNEY/AGENT INFORMATION:		
	NAME:	Brookes, A. Anders	
	REGISTRATION NUMBER:	36,373	
	REFERENCE/DOCKET NUMBER:	PF385	
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE:	301-309-8504	
	TELEFAX:	301-309-8439	
	INFORMATION FOR SEQ ID NO:	1:	
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	1710 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	linear	
	MOLECULE TYPE:	DNA (genomic)	
	FEATURE:		
	NAME/KEY:	CDS	
	LOCATION:	46..720	
	FEATURE:		
	NAME/KEY:	sig_peptide	
	LOCATION:	46..126	
	FEATURE:		
	NAME/KEY:	mat_peptide	
	LOCATION:	127..720	
	US-09-106-182-1		
	Query Match	94.0%; Score 749.4; DB 3; Length 1710;	
	Best Local Similarity	99.9%; Pred. No. 8.1e-170;	
	Matches	750; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	45	GCTTCGGGAGAGCGCACCCGCCGCGCAGCCCCCATGTGCACCTTCCGAGCA	104
DB	1	GCTTCGGGAGAGCGCACCCGCCGCGCAGCCCCCATGTGCACCTTCCGAGCA	60
QY	105	GGGACTCGTGCGGATTTAGCGTGCTGTGCACGGTGCTGTGCACCTTCCGAGTG	164
DB	61	GGGACTCGTGCGGATTTAGCGTGCTGTGCACGGTGCTGTGCACCTTCCGAGTG	120
QY	165	CAAGTCTTAATPGACAAGGACCAGCGCTTGGCCCCTCCATCCAGAATAAATATGAC	224
DB	121	CAAGTCTTAATPGACAAGGACCAGCGCTTGGCCCCTCCATCCAGAATAAATATGAC	180
QY	225	CTACCCCGCTACCTGGAGACCAACTCCGAGCTTGGCTGGGACCTATCTGAATACCTG	284
DB	181	CTACCCCGCTACCTGGAGACCAACTCCGAGCTTGGCTGGGACCTATCTGAATACCTG	240
QY	285	GGCCCCCTTTCAAAGCAGACTTCAACCTTCCCGCTGGGGCAGAGACTCTGCCC	344
DB	241	GGCCCCCTTTCAAAGCAGACTTCAACCTTCCCGCTGGGGCAGAGACTCTGCCC	300
QY	345	AGGGCACTGTGACTTGAGGTGTGGGAAAGCCTCAATGACAAAATCGGGCTGACCCAG	404
DB	301	AGGGCACTGTGACTTGAGGTGTGGGAAAGCCTCAATGACAAAATCGGGCTGACCCAG	360
QY	405	AATACAGGCGCTACAGCCACCTTCTGTGTACTTGGTGGCCTCAACCGTAGGCTGCC	464
DB	361	AATACAGGCGCTACAGCCACCTTCTGTGTACTTGGTGGCCTCAACCGTAGGCTGCC	420
QY	465	ACTGCTGAGCTGGCGCAGCTGCGCCACTTCTGACACCGCTCCAGGGCCTGCTGGGC	480
DB	421	ACTGCTGAGCTGGCGCAGCTGCGCCACTTCTGACACCGCTCCAGGGCCTGCTGGGC	480
QY	525	AGCATTCGGGGCGTCATGCGAGCTCTGGGCTAACCACTGCCCCAGCGCTGCTGGGACT	584
DB	481	AGCATTCGGGGCGTCATGCGAGCTCTGGGCTAACCACTGCCCCAGCGCTGCTGGGACT	540
QY	585	GAACCCACTTTGGACTCTCTGGCCCTGCCACACAGTGAATTTCTCCAGAAAGATGACGACTTC	644
DB	541	GAACCCACTTTGGACTCTCTGGCCCTGCCACACAGTGAATTTCTCCAGAAAGATGACGACTTC	600
QY	645	TGGCTGCTGAAGAGCTGCAGACCTGCTGCGCTGCGCCCAAGGACTTCAACCGGCTC	704
DB	601	TGGCTGCTGAAGAGCTGCAGACCTGCTGCGCTGCGCCCAAGGACTTCAACCGGCTC	660
QY	705	AAGAAGAAGATGCAGCCTCCAGCAGCTGACGTCAACCTTGACCTGGGGGCTCATGGCTTC	764
DB	661	AAGAAGAAGATGCAGCCTCCAGCAGCTGACGTCAACCTTGACCTGGGGGCTCATGGCTTC	720
QY	765	TGACTTCTGACCTTCTCTCTCTTCTGCTCCCCC	795
DB	721	TGACTTCTGACCTTCTCTCTCTTCTGCTCCCCC	751
	RESULT 5		
	US-08-792-019B-4		
	; Sequence 4, Application US/08792019B		
	; Patent No. 5741772		
	; GENERAL INFORMATION:		
	; APPLICANT: CHANG, MING-SHI		
	; TITLE OF INVENTION: THE NEUTROTROPIC FACTOR NNT-1		
	; NUMBER OF SEQUENCES: 16		
	; CORRESPONDENCE ADDRESS:		
	; ADDRESSEE: AMGEN INC.		
	; STREET: 1840 DEHAVILLAND DRIVE		
	; CITY: THOUSAND OAKS		
	; STATE: CA		
	; COUNTRY: USA		
	; ZIP: 91320		
	; COMPUTER READABLE FORM:		
	; MEDIUM TYPE: Floppy disk		
	; COMPUTER: IBM PC compatible		
	; OPERATING SYSTEM: PC-DOS/MS-DOS		
	; SOFTWARE: PatentIn Release #1.0, Version #1.30		
	; CURRENT APPLICATION DATA:		
	; APPLICATION NUMBER: US/08/792,019B		
	; FILING DATE: 03-FEB-1997		
	; CLASSIFICATION: 514		
	; ATTORNEY/AGENT INFORMATION:		
	; NAME: COOK, ROBERT R.		
	; REGISTRATION NUMBER: 31,602		
	; REFERENCE/DOCKET NUMBER: A-442		
	; INFORMATION FOR SEQ ID NO: 4:		
	; SEQUENCE CHARACTERISTICS:		
	; LENGTH: 819 base pairs		
	; TYPE: nucleic acid		
	; STRANDEDNESS: single		
	; TOPOLOGY: linear		
	; MOLECULE TYPE: CDNA		
	; FEATURE:		
	; NAME/KEY: CDS		
	; LOCATION: 95..769		
	; FEATURE:		
	; NAME/KEY: mat_peptide		
	; LOCATION: 176..769		
	; FEATURE:		
	; NAME/KEY: sig peptide		
	; LOCATION: 95..175		
</			

```

Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY 1 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTCGGAGAGGAG 60
Db 1 |||||
QY 5 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTGGGAGAGGAG 64
Db 5 |||||
QY 61 CGCAGCCCGCGCGCCAG-CCCCAGCCCATGAGACTCGAGCAGGGAGCTGTGGGG 119
Db 61 |||||
QY 65 CGCGCGCGCGCGCGCGCCAGCCCATGAGACTCGAGCAGGGAGCTGTGGGG 124
Db 65 |||||
QY 120 ATGTTAGCTGTGCTGACGCTGCTGCGACCTCGCCCTCGAGTGCAGCTCTCAATCG 179
Db 120 |||||
QY 125 ATGTTAGCTGTGCTATGACGCTGCTGCGACCTCGCCCTCGAGTGCAGCTCTTAATCG 184
Db 125 |||||
QY 180 ACAGGGACCCAGGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTG 239
Db 180 |||||
QY 185 ACAGGAGATCCAGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTG 244
Db 185 |||||
QY 240 GAGCACCACCTCGCGAGCTTGGCTGGGACCTATCTGAACTACCTGGGCCCCCTTTCAAC 299
Db 240 |||||
QY 245 GAGCATCAACTCGCGAGCTTAGCTGGGACCTACCTGAACTAAGCTGAACTCAAC 304
Db 245 |||||
QY 300 GAGCAGACTTCAACCTCCCGCTCGCGCGGCGAGACTCTGCCAGGGCCACTGTTGAC 359
Db 300 |||||
QY 305 GAGCCTGACTTCAATCTCTCGACTGGGCGAGAACTCTGCCAGGGCCACTGTTGAC 364
Db 305 |||||
QY 360 TTGAGGCTGTGGCGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTAC 419
Db 360 |||||
QY 365 TTGGAAGTGTGGCGAAGCTCAATGACAGGCTGGGCTGACCCAGAACTATGAGGCTAC 424
Db 365 |||||
QY 420 AGCCACCTTCTGTGTTACTTGGTGGGCTCAACCGTCAAGGCTCAAGAGAGATGAC 479
Db 420 |||||
QY 425 AGTCACTTCTGTGTTACTTGGTGGGCTCAACCGTCAAGGCTCAAGAGAGATGAC 484
Db 425 |||||
QY 480 CGCAGCTGGGCGGCTTCTGACAGGCTCGAGGCTCGCGGCTGACCCAGAACTACGAGGCTAC 539
Db 480 |||||
QY 485 CGTAGCTGGGCGGCTTCTGTAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCT 544
Db 485 |||||
QY 540 ATGGAGCTTCTGGCTTACCACTGCGCCAGCGCTCGCTGGGCTGAACTGAACTGGACT 599
Db 540 |||||
QY 545 ATGGCGAGCTTGGCTTACCACTGCGCCAGCTCTGCGAGGCTGAGGCTGGGCTGGG 604
Db 545 |||||
QY 600 CTGCGCTGCGCCAGAGTACTTCTCCAGAGATGAGGCTTCTGGCTGTGAAGGAG 659
Db 600 |||||
QY 605 CTGCGCTGCGCCAGAGTACTTCTCCAGAGATGAGGCTTCTGGCTGTGAAGGAG 664
Db 605 |||||
QY 660 CTGAGAGCTGCTGTGGGCTCGCGCAGGCTTCAACCGCTCAAGAGAGATGAC 719
Db 660 |||||
QY 665 CTGAGAGCTGCTGTGGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTGAC 724
Db 665 |||||
QY 720 CTTCCAGAGCTGAGTCACTCGCTGGGCTGAGGCTTCTGACTTCTGACTTCTGACTT 778
Db 720 |||||
QY 725 CTTCCAGAGCTTCACTGAGTCACTCGCTGGGCTGAGGCTTCTGACTTCTGACTTCTGACTT 783
Db 725 |||||

```

RESULT 6

```

US-08-988-819-4
; Sequence 4, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 176..769
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..175
; US-08-988-819-4

```

Query Match 84.0%; Score 669.4; DB 3; Length 819;
 Best Local Similarity 92.0%; Pred. No. 7.7e-151;
 Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

```

QY 1 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTCGGAGAGGAG 60
Db 1 |||||
QY 5 ATTAAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCGCCAGCTCTGGGAGAGGAG 64
Db 5 |||||
QY 61 CGCAGCCCGCGCGCCAG-CCCCAGCCCATGAGACTCTGCCAGGGCCACTGTTGAG 119
Db 61 |||||
QY 65 CGCGCGCGCGCGCGCGCCAGCCCATGAGACTCGAGTGCAGCTCTCAATCG 179
Db 65 |||||
QY 120 ATGTTAGCTGTGCTGACGCTGCTGCGACCTCGCGCTGACCCAGAACTACGAGGCTACCTG 239
Db 120 |||||
QY 125 ATGTTAGCTGTGCTATGACGCTGCTGCGACCTCGCCCTCGAGTGCAGCTCTTAATCG 184
Db 125 |||||
QY 180 ACAGGGACCCAGGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTG 244
Db 180 |||||
QY 185 ACAGGAGATCCAGGCTCGCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTG 299
Db 185 |||||
QY 240 GAGCACCACCTCGCGAGCTTGGCTGGGACCTATCTGAACTACCTGGGCCCCCTTTCAAC 304
Db 240 |||||
QY 245 GAGCATCAACTCGCGAGCTTAGCTGGGACCTACCTGAACTAAGCTGAACTCAAC 359
Db 245 |||||
QY 300 GAGCAGACTTCAACCTCCCGCTCGCGCGGCGAGACTCTGCCAGGGCCACTGTTGAC 364
Db 300 |||||
QY 305 GAGCCTGACTTCAATCTCTCGACTGGGCGAGAACTCTGCCAGGGCCACTGTTGAC 419
Db 305 |||||
QY 360 TTGAGGCTGTGGCGAAGCTCAATGACAACTCGCGCTGACCCAGAACTATGAGGCTAC 424
Db 360 |||||
QY 365 TTGGAAGTGTGGCGAAGCTCAATGACAGGCTGGGCTGACCCAGAACTATGAGGCTAC 479
Db 365 |||||
QY 420 AGCCACCTTCTGTGTTACTTGGTGGGCTCAACCGTCAAGGCTCAAGAGAGATGAC 484
Db 420 |||||
QY 425 AGTCACTTCTGTGTTACTTGGTGGGCTCAACCGTCAAGGCTCAAGAGAGATGAC 539
Db 425 |||||
QY 480 CGCAGCTGGGCGGCTTCTGACAGGCTCGAGGCTCGCGGCTGACCCAGAACTACGAGGCTAC 544
Db 480 |||||
QY 485 CGTAGCTGGGCGGCTTCTGTAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCT 599
Db 485 |||||
QY 540 ATGGAGCTTCTGGCTTACCACTGCGCCAGCGCTCGCTGGGCTGAACTGAACTGGACT 599
Db 540 |||||

```

Db 545 ATGGGAGCGCTTGGCTACCCAGTCCCGCCAGCCCTCTGCCAGGAGTGAAGCCAGCGCTGGGCC 604
Qy 600 CTGGCCCTGCCACAGTGAATCTCTCCAGAGATGAGAGACTTCTGGCTGCTGAAGGAG 659
Db 605 CTGGCCCTGCCACAGTGAATCTCTCCAGAGATGAGAGACTTCTGGCTGCTGAAGGAG 664
Qy 660 CTGACAGCTGGCTGGCTGGCCCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGAGATGAG 719
Db 665 CTGACAGCTGGCTGGCTGGCCCTGGCCCAAGGACTTCAACCGGCTCAAGAGAGATGAG 724
Qy 720 CTCCAGAGCTGAGTCAACCTGACCTGGGGCTCATGGCTTCTGACTTCTGACTT 778
Db 725 CTCCAGAGCTTCAAGTCAACCTGACCTGGAGGCAATGGTTTCTGACCTCTGACCT 783

RESULT 7

US-09-016-534-4
; Sequence 4, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 176..769
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 95..175
; US-09-016-534-4

Query Match 84.0%; Score 669.4; DB 3; Length 819;
Best Local Similarity 92.0%; Pred. No. 7.7e-151;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
Qy 1 ATTTAAAGCTTCGCGGAGCCCGGCTCGCCCTCCACCTCCGACGCTCCCGGAGAGGAG 60
|||||

Db 5 ATTTAAAGCTTCGCGGAGCCCGGCTCGCCCTCCACCTCCGACAGCTCTGGAGAGGAG 64
Qy 61 CCGCACCCGCGCGCCAG - CCGACGCCCCATGAGACCTCCGAGCAGGAGGACTCGTGGGG 119
Db 65 CCGCGCCGCGCGCGCGCGCCCGCCATGAGACCTCCGAGCAGGAGGACTCGTGGGG 124
Qy 120 ATGTTAGCGTCCCTGTGACCGGTCTGCGACCTCTCCCTGCGAGTGCAGCTCTCAATCGC 179
Db 125 ATGTTAGCTTCCCTATGACCGGTCTGTTGGCACCTCTCCCTGCGAGTGCAGCTCTTAATCGC 184
Qy 180 ACAGGGACCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACCTCAACCGCTACCTG 239
Db 185 ACAGGAGATCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACCTCAACCGCTACCTG 244
Qy 240 GAGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGCCCCCTTTCAAC 299
Db 245 GAGCATCAACTCCGAGCTTGGCTGGGACCTACTGAGTACTCTGGGCCCCCTTTCAAC 304
Qy 300 GAGCAGACTTTCAACCTCCCGCTGGGGGAGAGACTCTGCGCCAGGGCCACTGTTGAC 359
Db 305 GAGCTGACTTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGGCCACGCTCAAC 364
Qy 360 TTGAGGTGTGGCGAAGCTCAATGACAAACTGGCGCTGACCCAGAACTACGAGGGCTAC 419
Db 365 TTGGAAGTGTGGCGAAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGGCTAC 424
Qy 420 AGCCACCTTCTGTGTACTTGGCTGGCCCTCAACCGTCAAGCTGCGCTGAGCTGCGC 479
Db 425 AGTCACTTCTGTGTACTTGGCTGGCCCTCAACCGTCAAGCTGCGCTGAGCTGCGC 484
Qy 480 GCGAGCTTGGCCCACTTCTGACACGCTTCCAGGGCTCTGCGGAGAGCTTGGGGGCTC 539
Db 485 CGTAGCTTGGCCCACTTCTGTACTAGAGCTTCCAGGGCTCTGCGGAGAGCTTGGAGGTG 544
Qy 540 ATGCGAGCTTGGGCTACCACTGCGCCAGCGCTGCTGGGAGTGAACCCACTTGGACT 599
Db 545 ATGGCGAGCTTGGCTACCACTGCGCCAGCGCTGCTGGGAGTGAACCCACTTGGGCT 604
Qy 600 CTGGCCCTGCCACAGTGAATCTCTCCAGAAAGTGAAGAGTGAAGTGAAGGAG 659
Db 605 CTGGCCCTGCCACAGTGAATCTCTCCAGAAAGTGAAGTGAAGTGAAGGAG 664
Qy 660 CTGACAGCTGGCTGTGGCTGCGGCTCGGCGAAAGACTTCAACCGGCTCAAGAGAGATGAG 719
Db 665 CTGACAGCTGGCTGTGGCTTCAAGCAAGGACTTCAAGCGGCTTGAAGAGAGATGAG 724
Qy 720 CTCCAGAGCTGAGTCAACCTGACCTGCGGCTCATGGCTTCTGACTTCTGACCTT 778
Db 725 CTCCAGAGCTTCAAGTCAACCTGACCTTGGAGGAGATGAGTCTTCTGACCTT 783

RESULT 8

US-08-792-019B-3
; Sequence 3, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/792,019B

```
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"
; OTHER INFORMATION: REGION OF >1KB"
US-08-792-019B-3

Query Match 65.7%; Score 523.4; DB 1; Length 5087;
Best Local Similarity 99.8%; Pred. No. 6.9e-116;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACTGGGCCCCCTTTCAACAGGAGCAGACTTCAACCTCCCGCCTGGGG 330
Db 3363 AGCTGAACCTACTGGGCCCCCTTTCAACAGGAGCAGACTTCAACCTCCCGCCTGGGG 3422

QY 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGGAAGCCTCAATGACAAAC 390
Db 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGGAAGCCTCAATGACAAAC 3482

QY 391 TCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGCTGCTCA 450
Db 3483 TCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGCTGCTCA 450

QY 451 ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCCTGCGCCACTTCTGACAGCCTCC 510
Db 3543 ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCCTGCGCCACTTCTGACAGCCTCC 3602

QY 511 AGGGCTCTGTCGGGAGCATTGGGGCGTTCATGGAGCTGCGGCTACCCACTGCCCCAGC 570
Db 3603 AGGGCTCTGTCGGGAGCATTGGGGCGTTCATGGAGCTGCGGCTACCCACTGCCCCAGC 3662

QY 571 CGCTGCTGGGACTGAACCCACTTGGAGCTCTGGGCTGCCCCAGAGTGAACCTTCTCC 630
Db 3663 CGCTGCTGGGACTGAACCCACTTGGAGCTCTGGGCTGCCCCAGAGTGAACCTTCTCC 3722

QY 631 AGATGAGCAGCTTCTGGCTGCTGAAGAGCTGCGAGCCTGCGAGCTGCGAGCTGCGAG 690
Db 3723 AGATGAGCAGCTTCTGGCTGCTGAAGAGCTGCGAGCCTGCGAGCTGCGAGCTGCGAG 3782

QY 691 ACTTCAACCGGCTCAAGAGAGATGCGAGCTTCTGCTTCTGCTGCTCCCC 750
Db 3783 ACTTCAACCGGCTCAAGAGAGATGCGAGCTTCTGCTTCTGCTGCTCCCC 3842

QY 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCC 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCC 3887

RESULT 9
US-08-988-819-3
; Sequence 3, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
```

```
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"
; OTHER INFORMATION: REGION OF >1KB"
US-08-988-819-3

Query Match 65.7%; Score 523.4; DB 3; Length 5087;
Best Local Similarity 99.8%; Pred. No. 6.9e-116;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 271 ATCTGAACCTACTGGGCCCCCTTTCAACAGGAGCAGACTTCAACCTCCCGCCTGGGG 330
Db 3363 AGCTGAACCTACTGGGCCCCCTTTCAACAGGAGCAGACTTCAACCTCCCGCCTGGGG 3422

QY 331 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGGAAGCCTCAATGACAAAC 390
Db 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGGGAAGCCTCAATGACAAAC 3482

QY 391 TCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGCTGCTCA 450
Db 3483 TCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGCTGCTCA 3542

QY 451 ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCCTGCGCCACTTCTGACAGCCTCC 510
Db 3543 ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCCTGCGCCACTTCTGACAGCCTCC 3602

QY 511 AGGGCTCTGTCGGGAGCATTGGGGCGTTCATGGAGCTGCGGCTACCCACTGCCCCAGC 570
Db 3603 AGGGCTCTGTCGGGAGCATTGGGGCGTTCATGGAGCTGCGGCTACCCACTGCCCCAGC 3662

QY 571 CGCTGCTGGGACTGAACCCACTTGGAGCTCTGGGCTGCCCCAGAGTGAACCTTCTCC 630
Db 3663 CGCTGCTGGGACTGAACCCACTTGGAGCTCTGGGCTGCCCCAGAGTGAACCTTCTCC 3722

QY 631 AGATGAGCAGCTTCTGGCTGCTGAAGAGCTGCGAGCCTGCGAGCTGCGAGCTGCGAG 690
Db 3723 AGATGAGCAGCTTCTGGCTGCTGAAGAGCTGCGAGCCTGCGAGCTGCGAGCTGCGAG 3782

QY 691 ACTTCAACCGGCTCAAGAGAGATGCGAGCTTCTGCTTCTGCTGCTCCCC 750
Db 3783 ACTTCAACCGGCTCAAGAGAGATGCGAGCTTCTGCTTCTGCTGCTCCCC 3842

QY 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCC 795
Db 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTCCCC 3887
```

RESULT 10
US-09-016-534-3
; Sequence 3, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"
; OTHER INFORMATION: REGION OF >1KB"
US-09-016-534-3

Query Match 65.7%; Score 523.4; DB 3; Length 5087;
Best Local Similarity 99.8%; Pred. No. 6.9e-116;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 271 ATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 330
DB 3363 AGCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 3422
QY 331 CAGAGACTCTGCCGAGGCGCACTTGTGAGTGGAGGTGGGAGGCTCAATGACAAAC 390
DB 3423 CAGAGACTCTGCCGAGGCGCACTTGTGAGTGGAGGTGGGAGGCTCAATGACAAAC 3482
QY 391 TGGGGCTGACCCAGAACTACGAGGCGCTACAGCCACCTTCTGTGTACTTGGGTGGCTCA 450
DB 3483 TGGGGCTGACCCAGAACTACGAGGCGCTACAGCCACCTTCTGTGTACTTGGGTGGCTCA 3542
QY 451 ACCGTAGGTGCGCACTGTGAGTGGGCGGAGCGCTGCGGCGCACTTCTGTGACAGCCTCC 510
DB 3543 ACCGTAGGTGCGCACTGTGAGTGGGCGGAGCGCTGCGGCGCACTTCTGTGACAGCCTCC 3602
QY 511 AGGCGCTGCTGGGCGAGCAATTGGGCGGCGTATGCGAGCTCTGGGCTACCACTGCCCCAGC 570
DB 3603 AGGCGCTGCTGGGCGAGCAATTGGGCGGCGTATGCGAGCTCTGGGCTACCACTGCCCCAGC 3662

QY 571 CGTGCCTGGGACTGAACCCACTTGGACTCCTGGCCCTGCCCAAGTGAAGTCTTCTCCAG 630
DB 3663 CGTGCCTGGGACTGAACCCACTTGGACTCCTGGCCCTGCCCAAGTGAAGTCTTCTCCAG 3722
QY 631 AGATGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGCTGCTGGCTGCGGCAAGG 690
DB 3723 AGATGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGCTGCTGGCTGCGGCAAGG 3782
QY 691 ACTTCAACCGGCTCAAGAAGATGACAGCTTCCAGCAGCTGCAGTGCAGTCAACCTGCACCTGG 750
DB 3783 ACTTCAACCGGCTCAAGAAGATGACAGCTTCCAGCAGCTGCAGTGCAGTCAACCTGCACCTGG 3842
QY 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTGACCTTCTCTCTTCGCTCCCTCC 795
DB 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTCGCTCCCTCC 3887

RESULT 11
US-09-106-182-7
; Sequence 7, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-106-182-7

Query Match 10.1%; Score 80.2; DB 3; Length 396;
Best Local Similarity 96.5%; Pred. No. 1.4e-10;
Matches 82; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 711 AAGATGACGCTTCCAGCAGCTGCAGTCAACCTGCAGCTGGGGCTCATGGCTTCTGACTT 770
DB 4 ACGAGGCGAGCTTCCAGCAGCTGCAGTCAACCTGCAGCTGGGGCTCATGGCTTCTGACTT 63
QY 771 CTGACCTTCTCTCTTCGCTCCCTCC 795
DB 64 CTGACCTTCTCTCTTCGCTCCCTCC 88

RESULT 12

US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 5.6%; Score 44.8; DB 3; Length 68750;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 112; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 456 CAGGCTGCCACTGCTGAGCTGCGCGGAGCTGCGCCACCTTCTGCACAGCCTCCAGGGC 515
DB 57992 CAGGTAGCGAGTCGGGAGCGCGGGGAAACGTGCGCGCCCTTCCGGATCCAGCCCGCGGC 58051

QY 516 CTGCTGGGAGCATTCGGCGGTCATGCGAGCTCTGGGCTACCCACTGCCCCAGCCGCTG 575
DB 58052 GTGCTGGCGGATCGGGCGCTCGAGACCCCGCGAGGTGGAACCTGCGCCGCTGCGC 58111

QY 576 CTTGGGACTGAACCCACTTGGACTCTTGGGCTACCCACTGCCCCAGCCGCTG 575
DB 58112 GCCCGGCTTCATGCGCGCTGCGCCCTGCGGCTATCTATGCTGCGCTCGCCGAGATGGG 58171

QY 636 GACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGCTGCTGGCG 679
DB 58172 CTTCAATACGGCGCGCTTGGCGGGGCTCGCCGAGCTGTGGCG 58215

RESULT 13

US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 5.6%; Score 44.8; DB 4; Length 68750;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 112; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 456 CAGGCTGCCACTGCTGAGCTGCGCGGAGCTGCGCCACCTTCTGCACAGCCTCCAGGGC 515
DB 57992 CAGGTAGCGAGTCGGGAGCGCGGGGAAACGTGCGCGCCCTTCCGGATCCAGCCCGCGGC 58051

QY 516 CTGCTGGGAGCATTCGGCGGTCATGCGAGCTCTGGGCTACCCACTGCCCCAGCCGCTG 575
DB 58052 GTGCTGGCGGATCGGGCGCTCGAGACCCCGCGAGGTGGAACCTGCGCCGCTGCGC 58111

QY 576 CTTGGGACTGAACCCACTTGGACTCTTGGGCTACCCACTGCCCCAGCCGCTG 635
DB 58112 GCCCGGCTTCATGCGCGCTGCGCCCTGCGGCTATCTATGCTGCGCTCGCCGAGATGGG 58171

QY 636 GACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGCTGCTGGCG 679
DB 58172 CTTCAATACGGCGCGCTTGGCGGGGCTCGCCGAGCTGTGGCG 58215

RESULT 14

US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 5.6%; Score 44.8; DB 4; Length 68750;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 112; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 456 CAGGCTGCCACTGCTGAGCTGCGCGGAGCTGCGCCACCTTCTGCACAGCCTCCAGGGC 515
DB 57992 CAGGTAGCGAGTCGGGAGCGCGGGGAAACGTGCGCGCCCTTCCGGATCCAGCCCGCGGC 58051

QY 516 CTGCTGGGAGCATTCGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAGCCGCTG 575
DB 58052 GTGCTGGCGGATCGGGCGCTCGAGACCCCGGAGGTGGAACCTGCGCCGCTGCGC 58111

QY 576 CTTGGGACTGAACCCACTTGGACTCTTGGCTACCCACTGCCCCAGCCGCTG 635
DB 58112 GCCCGGCTTCATGCGCGCTGCGCCCTGCGGCTATCTATGCTGCGCTCGCCGAGATGGG 58171

QY 636 GACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGCTGCTGGCG 679
DB 58172 CTTCAATACGGCGCGCTTGGCGGGGCTCGCCGAGCTGTGGCG 58215

RESULT 15

US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross

us-09-931-704-1.rni

Sun Feb 2 08:31:38 2003

```

; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

```

```

Query Match      5.6%; Score 44.8; DB 4; Length 68750;
Best Local Similarity 50.0%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 112;
QY 456 CAGGCTGCCACTGCTGAGCTGCCCGCAGCTGGCCCACTTCTGCACCCAGCTCCAGGGC 515
Db 57992 CAGGTAGCGAGCTCGGAGCCCGGGGAGAGCTCGCGGCCCTTCGGATCCACGCCCGGGC 58051
QY 516 CTGCTGGGCAGCAATGCGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCAGCCAGCCCTG 575
Db 58052 GTGCTGCCCGGATCGGGCGCGTTCGAGACCCCGGCGAGGTGAACTCGCCGCCCTGCGC 58111
QY 576 CTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTACTTCTCCAGAAAGATG 635
Db 58112 GCCCGGCTTCATGCCCGCTGCCCGCTCGCGCTATCTATGTGGCTCGCGAGATGGGG 58171
QY 636 GACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGGCTGTGGCG 679
Db 58172 CTTCAATACGCCCGCGCTTCGGGGGGCTCGCGAGCTGTGGCG 58215

```

Search completed: February 1, 2003, 08:46:09
Job time : 240.683 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:22 ; Search time 14.5 Seconds
(without alignments)
1491.741 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGLACTLVLM.....KKMKQPPAAAVTLHLGAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	124.5	10.2	203	2 JC4645	cardiotrophin-1 -
2	118.5	9.7	203	2 I49153	cardiotrophin-1 -
3	92	7.5	332	2 G86182	hypothetical prote
4	91.5	7.5	201	2 G02312	cardiotrophin-1 -
5	91.5	7.5	1182	2 I48378	hairless protein -
6	91	7.4	195	2 JH0680	ciliary neurotroph
7	89.5	7.3	1313	2 T38943	probable guanine n
8	89	7.3	733	2 D83588	conserved hypotet
9	87	7.1	955	2 T00247	zinc finger protei
10	87	7.1	1561	2 T00248	zinc finger protei
11	86.5	7.1	640	2 T34916	transferase - stre
12	85.5	7.0	560	2 C38604	poly(3-hydroxyalka
13	84.5	6.9	389	2 F72511	probable cystathio
14	84	6.9	200	2 AD3633	hypothetical prote
15	84	6.9	411	2 F75439	probable cell wall
16	84	6.9	559	2 G71327	probable apolipop
17	83	6.8	452	2 A35602	tailless (tbl) pro
18	82.5	6.7	542	2 A82965	hypothetical prote
19	82.5	6.7	723	2 B38749	3-phosphatidylinos
20	82	6.7	1179	2 T04584	TWV resistance pro
21	81	6.6	479	2 C97508	blue-light photore
22	81	6.6	479	2 A42726	DNA photolase [im
23	80.5	6.6	353	2 AB1823	hypothetical prote
24	80.5	6.6	512	2 S21171	actinin receptor S
25	80	6.5	200	1 UNRTCF	ciliary neurotroph
26	80	6.5	460	2 D75493	cell division cycl
27	80	6.5	2261	1 A42548	genome polyprotein
28	79	6.4	812	2 T34180	hypothetical prote
29	79	6.4	1220	2 AD0125	exodeoxyribonuclea

protein-tyrosine k
hypothetical prote
conserved hypotet
yes-associated pro
hypothetical prote
CDA peptide synthe
hypothetical prote
hypothetical prote
extracellular solu
65k early nonstruc
65k early nonstruc
hypothetical prote
conserved hypotet
hypothetical prote
conserved hypotet

30 79 6.4 1621 2 T30200
31 78.5 6.4 390 2 T35509
32 78.5 6.4 426 2 A10287
33 78 6.4 254 2 T08755
34 77.5 6.4 272 2 T35231
35 77.5 6.3 741 2 A83271
36 77.5 6.3 7463 2 T36248
37 77 6.3 322 2 T23891
38 77 6.3 344 2 T15457
39 77 6.3 520 2 B75332
40 77 6.3 586 1 WMBEDE
41 77 6.3 587 1 WMBETE
42 77 6.3 733 2 A82765
43 76.5 6.2 310 2 A82734
44 76.5 6.2 338 2 T45394
45 76.5 6.2 522 2 D87123

ALIGNMENTS

RESULT 1

JC4645

cardiotrophin-1 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000

C;Accession: JC4645

R;Ishikawa, M.; Saito, Y.; Miyamoto, Y.; Kuwahara, K.; Ogawa, E.; Nakagawa, O.; Harada

Biochem. Biophys. Res. Commun. 219, 377-381, 1996

A;Title: CDNA cloning of rat cardiotrophin-1 (CT-1): Augmented expression of CT-1 gene

A;Reference number: JC4645; MUID:96193659; PMID:8604995

A;Accession: JC4645

A;Molecule type: mRNA

A;Residues: 1-203 <ISH>

A;Cross-references: DDBJ:D78591; NID:gl256926; PIDN:BAA11427.1; PID:gl256927

C;Genetics:

A;Gene: CT-1

C;Keywords: cardiac muscle; cytokine; heart

Query Match 10.2%; Score 124.5; DB 2; Length 203;

Best Local Similarity 28.1%; Pred. No. 0.00045;

Matches 50; Conservative 24; Mismatches 85; Indels 19; Gaps 5;

QY 40 IQKTYDLTRVLEHQRLSLAGTYLNLGPFNEPDPFNPRL---GAETLPRAVTDLVWRS 96

Db 27 IRQTHNLARLLTKYADQLLEEVVQQQEPFGLPGFSPRLPLAGLSGPAPSHAGLPV--- 83

QY 97 LNDKLRITQNYEAYSHLLCYLRGLNRQAA-----TAE LRSLAHFCTSLQGLLSIAGVM 151

Db 84 ---SERLRQDAALSAIPALIDAVRRRQALNPAPRLRLSLRSLDAAARQVRAAGAAVTVL 140

QY 152 AALGY----PLPQPLPGTEPTWTGPAHSDFLQKMDDFWLLKELQTLWLRSAKDPNRL 205

Db 141 AALGAAARGVPPEPV-ATSAFTSNAAGVFSKVLGLHVCGLYGEWVSRTEGDLGQL 197

RESULT 2

I49153

cardiotrophin-1 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C;Accession: I49153

R;Pennica, D.; King, K.L.; Shaw, K.J.; Luis, E.; Rullamas, J.; Luoh, S.; Darbonne, W.C

Proc. Natl. Acad. Sci. U.S.A. 92, 1142-1146, 1995

A;Title: Expression cloning of cardiotrophin 1, a cytokine that induces cardiac myocyte

A;Reference number: I49153; MUID:95166785; PMID:7862649

A;Accession: I49153

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-203 <RES>

A;Cross-references: EMBL:U18366; NID:g710331; PIDN:AAC52173.1; PID:g710332

C;Genetics:

A;Gene: ctfl

RESULTS 9
JH0680
ciliary neurotrophic factor - chicken

N;Alternate names: growth-promoting activity protein
C;Species: Gallus gallus (chicken)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000

C;Accession: J06080; P00057
R;Leung, D.W.; Parent, A.S.; Cachianes, G.; Esch, F.; Coulombe, J.N.; Nikolic, K.; Eickbush, H.; Scher, C.D.; 1045-1053, 1992

A;Title: Cloning, expression during development, and evidence for release of a trophic factor from embryonic periphery of chick retinal ganglion cells
A;Reference number: J06080; MUID:92304573; PMID:1610564

A;Accession: J06080

A;Molecule type: mRNA

A;Residues: 1-195 <LEU>

A;Cross-references: GB:M80827; NID:g211822; PID:AAA48784.1; PID:g211823

A;Experimental source: eye

R;Eckstein, F.P.; Esch, F.; Holbert, T.; Blacher, R.W.; Nishi, R.

Neuron 4, 623-631, 1990

A;Title: Purification and characterization of a trophic factor for embryonic peripheral

A;Reference number: P00057; MUID:90211978; PMID:2322465

A;Accession: P00057

A;Molecule type: protein

A;Residues: 155-166, 'X', 168-175 <ECK>

A;Experimental source: sciatic nerves

C;Comment: This is a neurotrophic protein.

C;Superfamily: ciliary neurotrophic factor

C;Keywords: growth factor

Query Match 7.4%; Score 91; DB 2; Length 195;
Best Local Similarity 27.3%; Pred. No. 0.53;
Matches 51; Conservative 21; Mismatches 83; Indels 32; Gaps 9;

QY 46 LTRYLEHQLRSAGTYLNYLGGPNEDFPNPRGATLPRATVDLEVRSLNDKRLQ 105

DB 23 LARKMRSDVTDLLIDYVERQG-----LDASISVAADVGVPTAAV--ERWAEQTGTORLLD 75

QY 106 N---YEAYSHLLCYLGLNQA---ATAELRSLA-----HPTSLQGLGSIAGVMA 152

DB 76 NLAAAYRFRLLAQMLEEQRELGGDTDAELGPALAAMLLQVSAPFYHLELL-----ELE 130

QY 153 ALGYPLPQPLPGTPTTPGPAH--SDFLQKMDDFWLLKELQTLWLRSAKDFNRLKKMQP 211

DB 131 SRGAPAE--GSEP---PAPPLSLFEQKLRLGLRLVRLAQWAVRSVRDLRLQSKGPG 184

QY 212 PAAAVTL 218

DB 185 SGAALGL 191

RESULT 7

T38943
probable guanine nucleotide binding protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T38943

R;Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, May 1997

A;Reference number: Z21819

A;Accession: T38943

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1313 <SKE>

A;Cross-references: EMBL:Z95396; PID:CA808769.1; GSPDB:GN00066; SPDB:SPAC57A7.11

A;Experimental source: strain 972h-; cosmid c57A7

C;Genetics:

A;Gene: SPDB:SPAC57A7.11

A;Map position: 1

Query Match 7.3%; Score 89.5; DB 2; Length 1313;
Best Local Similarity 22.8%; Pred. No. 8;
Matches 61; Conservative 32; Mismatches 69; Indels 105; Gaps 14;

QY 15 LCTVLMHLPV--PALNR-----TGDPGPGPSI-----QKTYDLTRYLEHQLRS 56

DB 811 LAFLLQHLPALHAKSLSKDTNTSVTSDPKPHFPVPSVSENKILNRSFSLTRSLKGLALS 870

QY 57 LAG-----TYLNYL-----GPP-FNEPDPNPPRLGATLPRATVDL 91
DB 871 LAGSDRASELLSNGENKPAESNLHLTSAKVPGPAPFALEY-----QSELD 919
QY 92 EVWRSND-----KLRLTQNYEAYSHLLCYLRLG---NRQATAELRSLAHECTS 139
DB 920 PLTSYLFDMRSKYTFTEQMPNEDDEPGS--ICYNQRLWRNRNEKLIYTRPLAEYSTN 977
QY 140 -----LQGLLSIAGVMAALGYPLPQPLPGTPTTPGPAHSDFLQKMDDFWLLKE--- 190
DB 978 GRNQQLMTFNNTIA-----PRKLMFHQFEDQLITLGDKI 1013
QY 191 LQTLWLRSAKDFNRLKKMQPPAAAVT 217
DB 1014 IQVWDRR-----NRCLNSFKTSASAT 1036

RESULT 8

DB3588

conserved hypothetical protein PA0454 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: D83588

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: D83588

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-733 <STO>

A;Cross-references: GB:AE004483; GB:AE004091; NID:g9946313; PIDN:AAG03843.1; GSPDB:GN0

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0454

C;Superfamily: hypothetical protein H11680

Query Match 7.3%; Score 89; DB 2; Length 733;

Best Local Similarity 26.8%; Pred. No. 4.3;

Matches 58; Conservative 26; Mismatches 66; Indels 68; Gaps 13;

QY 28 LNRTGDPGPGPSIQKTYDLTRYLE-----HOLRLAGTYLNYLGGPNEDP--FNPP 77

DB 230 LNRLGHRPGPKV-----SRYLKLYFIADQVHERASSSHYPNRLAEAFHSDVLRQC 283

QY 78 RL-----GATLPRATVDLEVRSLNDKRLTQNYEAYSHLLCYLRLGRLQAATAELRR 131

DB 284 RLNLQOGKACQALARA-----IRLRQPPD-YA-----DRELALEDLQA 320

QY 132 SLAHF-----CTSLOGLLSIAGVMAALGYPLPQPLPGTPTTPGPAHSDFLQKMDDFWL 187

DB 321 SLEHLRQOSNPAPWKGLRLSL-GALAAANLTTLDRLKLAGA-----SNPDIADEQDSAL 371

QY 188 LKELQTLWLRSAKD-FNRLKKMQPPAAAVTLHLGAGH 224

DB 372 LDRSP-----RSLKDAFERLRQQLTP-----TSLLFRHG 400

RESULT 9

T00247

zinc finger protein wiz - mouse

C;Species: Mus musculus (house mouse)

C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999

C;Accession: T00247

R;Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tomyama, M.

submitted to the EMBL Data Library, March 1998

A;Description: Molecular cloning and distinct developmental expression pattern of spli

A;Reference number: Z14130

A;Accession: T00247

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-955 <MAT>
A;Cross-references: EMBL:AB012266; NID:d1227741; PIDN:BAA32791.1; PID:d1033757
A;Experimental source: brain
C;Genetics: wiz
A;Gene: wiz

Query Match 7.1%; Score 87; DB 2; Length 955;
Best Local Similarity 27.8%; Pred. No. 9.2;
Matches 35; Conservative 19; Mismatches 48; Indels 24; Gaps 7;

QY 22 LPAVPAALNRTGPGPGPSIQKTYDLYLRYLHQLRSLAG---TYNLVILGP-----PFNEPD 73

Db 615 LFLSPLASRPGKPGAGPT-----QVPR--ELSLSPITGSKPSAASVILGPVATKRPQEDR 667

QY 74 FNPRLGAEITLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL---NROAATAELR 130

Db 668 FLPAEVKAKTYIQTELPPFKA-KTLHEK-----TSHSSTEACCELCGLYFENRKALASHAR 721

QY 131 RSLAHF 136

Db 722 AHLROF 727

RESULT 10

T00248

zinc finger protein wizL - mouse

C;Species: Mus musculus (house mouse)

C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999

C;Accession: T00248

R;Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiozaka, S.; Wanaka, A.; Tohyama, M.

submitted to the EMBL Data Library, March 1998

A;Description: Molecular cloning and distinct developmental expression pattern of splice

A;Reference number: Z14130

A;Accession: T00248

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1561 <MAT>

A;Cross-references: EMBL:AB012265; NID:d1227740; PIDN:BAA32790.1; PID:d1033756

A;Experimental source: brain

C;Genetics: wiz

A;Gene: wiz

Query Match 7.1%; Score 87; DB 2; Length 1561;
Best Local Similarity 27.8%; Pred. No. 17;
Matches 35; Conservative 19; Mismatches 48; Indels 24; Gaps 7;

QY 22 LPAVPAALNRTGPGPGPSIQKTYDLYLRYLHQLRSLAG---TYNLVILGP-----PFNEPD 73

Db 1221 LFLSPLASRPGKPGAGPT-----QVPR--ELSLSPITGSKPSAASVILGPVATKRPQEDR 1273

QY 74 FNPRLGAEITLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGL---NROAATAELR 130

Db 1274 FLPAEVKAKTYIQTELPPFKA-KTLHEK-----TSHSSTEACCELCGLYFENRKALASHAR 1327

QY 131 RSLAHF 136

Db 1328 AHLROF 1333

RESULT 11

T34916

transferase - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 02-Sep-2000

C;Accession: T34916

R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, January 1998

A;Reference number: Z21558

A;Accession: T34916

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-640

A;Cross-references: EMBL:AL021409; PIDN:CAA16181.1; GSPDB:GN00070; SCOEDB:SC3F7.10

A;Experimental source: strain A3 (2)

C;Genetics:

A;Gene: SCOEDB:SC3F7.10

C;Superfamily: glycine C-acetyltransferase homology

F;287-624/Domain: glycine C-acetyltransferase homology <GCA>

Query Match 7.1%; Score 86.5; DB 2; Length 640;
Best Local Similarity 27.2%; Pred. No. 6.2;
Matches 50; Conservative 13; Mismatches 64; Indels 57; Gaps 9;

QY 22 LPAVPAALNRTGPGPGPSIQKTYDLYLRYLHQL---RSLAGTYNLVILGPPFFNEPDENPR 78

Db 100 LPAVPA-EGTGDPTVEAVAAAMRQTQYRHLGLDADLEG-----E 140

QY 79 LGABTLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGLNRQAATAELRRSLAHFCT 138

Db 141 LGVDSVVITSVAAE---TERLGLT-----GAAPDAAGATTIRALA--D 179

QY 139 SLOGLLSIAGVMAALGYPLPQPLGTEPTWTGPAHSDFLOKMDDFWLLKELQTLWLR 198

Db 180 ALRGL-----VAAAAGTAVPEAAPATGAA-APAPGRSGNAP-----APGADGWDHRS 225

QY 199 AKDF 202

Db 226 MKDF 229

RESULT 12

C38604

poly(3-hydroxyalkanoate) polymerase (EC 2.7.7.-) 3 - Pseudomonas oleovorans

C;Species: Pseudomonas oleovorans

C;Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 08-Sep-2000

C;Accession: C38604

R;Huisman, G.W.; Wouink, E.; Meima, R.; Kazemier, B.; Terpstra, P.; Witholt, B.

J. Biol. Chem. 266, 2191-2198, 1991

A;Title: Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas oleovorans.

A;Reference number: A38604; MUID:91115830; PMID:1989978

A;Accession: C38604

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-560 <HUI>

A;Cross-references: GB:M58445; NID:G151441; PIDN:AAA25934.1; PID:G151444

C;Superfamily: poly(3-hydroxyalkanoic acid) synthase phbC

C;Keywords: nucleotidyltransferase

Query Match 7.0%; Score 85.5; DB 2; Length 560;
Best Local Similarity 30.2%; Pred. No. 6.4;
Matches 38; Conservative 17; Mismatches 42; Indels 29; Gaps 7;

QY 77 PRGAEITLPRATVDLEWVRSNDKRLRTQNYEAYSHLLCYLRGLNRQAATAELRRSL--A 134

Db 5 PAKGTPTLPATSMNVQ-----NAILGLRGR-----DLISLRNVSRQS---LRHPLHTA 50

QY 135 HFCTSLGLOGSIAGVMAALGYPLPQPLG-----TEPTWTGPAHSDFLOKMDDFWLLKE 190

Db 51 HHLALGGLGRV-----ILGDTPLQPNRPDRPFSPTWSQNPFFYRRGLQA----YLAQW 101

QY 191 LQTLW 196

Db 102 KQTRLW 107

RESULT 13

F72511

probable cystathionine gamma-synthase APE2068 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C;Accession: F72511

R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A;Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: F72511
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <RAW>
A:Cross-references: DDBJ:AF000063; NID:g5105654; PIDN:BAA81078.1; PID:g5105766
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2068
C:Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 6.9%; Score 84.5; DB 2; Length 389;
Best Local Similarity 22.9%; Pred. No. 5;
Matches 47; Conservative 19; Mismatches 78; Indels 61; Gaps 8;

QY 43 TYDLTRYLEHQLSLAGTYLNYLGGPPNEPDPNPPRLGAETLPRATVDLEWRSLSL----- 97
DB 107 TYGSTRSLEMLSSITGIEVRLAGPPWED-----LGLVCWADLIIVES 150
QY 98 --NDKRLRTQNYEAYSHLLCYLRGLNRQAATAELR-----RSL---AHFCTS 139
DB 151 MANPTLRVPP-----LSGIYREAGSCGVRVVVDNTFATPIAYRPLERGAHY--\$ 197
QY 140 LQGLLSGIAGVMAALGYPLPOPL-PGTEPTWPGPAHSDFLQKMDDFWLLKELQTLWLWRS 198
DB 198 LESLTKYIAGHNDVVGSLSGRVEEDLEPLNMWKILGTIMQPIDAY-----LAWRG 249
QY 199 AKDFNRLKKKQWPPAAAVTLHLGAH 223
DB 250 MKTLKARFEAQSAVAEVAEWLESH 274

RESULT 14

AD3633
hypothetical protein BMEI10989 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3633
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Los, T.; Ivanova, M.; Mazur, M.; Goldstein, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3633
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54231.1; PID:g17985203; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10989
A:Map position: II

Query Match 6.9%; Score 84; DB 2; Length 200;
Best Local Similarity 24.3%; Pred. No. 2.4;
Matches 56; Conservative 22; Mismatches 84; Indels 68; Gaps 13;

QY 10 GMLACLCTVWLHPALNRTGDPGPGPSIQKTYDTRYLEHQLSLAGTYLNLGLPFP 69
DB 2 GLLAGALV---LPSLPAEAKTQQAAPPNATSPHQADVYL---LRGFADIFSTGI---- 51
QY 70 NEPDFNPPRLGAETLPRATVDLEY-----WRSNDKRLRTQNYEAYSHLLCYLRG--LNR 122
DB 52 -----DEIGAE-LQAAGVNAHQHAAWELVNL-RIVADQKNGHLPVVLIGHSLGA 101
QY 123 QAA---TAEILRSIAHFTCSLQGLLSAGVMAALGYPLPOPLPGT-----BFTW 169
DB 102 NAAIYIAELER-----RGIAVDYMATFAATG---PDPLFCNVRRVNVFYFKQHW 149
QY 170 ---TPGPAHSDFLQKMDDFWLLKELQTLWLWRSKDFNRLKKKQWPPAAA 215
DB 150 GLPLVPGPRFHGHLENRD-----FSNAKDVGHFNIEKQRPLOA 187

RESULT 15

F75439
probable cell wall synthesis protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75439
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A/5250; MUID:20036896; PMID:10567266
A:Accession: F75439
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <WHI>
A:Cross-references: GB:AE001958; GB:AE000513; NID:g6458805; PIDN:AAF10649.1; PID:g6458805
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1076
A:Map position: 1

Query Match 6.9%; Score 84; DB 2; Length 411;
Best Local Similarity 24.6%; Pred. No. 6;
Matches 56; Conservative 25; Mismatches 77; Indels 70; Gaps 11;

QY 2 DLKAGDSWGLACLCTVWLHPALNRTGDPGPGPSIQKTYDTRYLEHQLSLAGTY 61
DB 212 EVLAADRWAELFSEMSLRPGVPLLLSGG-GRGHYAAADVLT-----ELGNL-GRA 264
QY 62 LNYLGGPPNEPDPNPPRLGAETLPRATVDLEWRSNDKRLRTQNYEAYSHLLCYLRGLN 121
DB 265 VQVLVPASRQGE-----GTETIGGATV-----HHLGFRRDLP 296
QY 122 RQATAEILRSIAHFTCSLQGLLSAGV-----MAALGYPLP-QPLPGTEPTWTPGPAH 175
DB 297 RLAAASDL-----VVKGAGGLTVAEATALGVPLVIYAPIPQGE-----EHN 337
QY 176 SDFLQKMDDFWLLKELQTLWLW-RSADFNRLKKKQWPPAAAATLHLGA 222
DB 338 ADFLERHG-----AGLWARHHDVRLVLRALDPAEHAHLSAGA 376

Search completed: January 27, 2003, 15:42:34
Job time : 16.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:22 ; Search time 9 Seconds
(without alignments)
1036.908 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGMCLCTVLW.....KKKMQPPAAAATLHLGAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	124.5	10.2	203	1	CTF1_RAT
2	118.5	9.7	203	1	CTF1_MOUSE
3	94	7.7	423	1	CST_MOUSE
4	93	7.6	619	1	NXF1_HUMAN
5	91.5	7.5	201	1	CTF1_HUMAN
6	91.5	7.5	1182	1	HAIR_MOUSE
7	91	7.4	195	1	CNTF_CHICK
8	89.5	7.3	1313	1	MIP1_SCHPO
9	87.5	7.1	1181	1	HAIR_RAT
10	86	7.0	1009	1	M2B2_HUMAN
11	85.5	7.0	560	1	PHAC_PSEOL
12	84	6.9	200	1	CNTF_PIG
13	83	6.8	452	1	TLL_DROME
14	82.5	6.7	1189	1	HAIR_HUMAN
15	82	6.7	724	1	F8B5_BOVIN
16	81	6.6	291	1	Y332_MYCTU
17	81	6.6	618	1	NXF1_RAT
18	80.5	6.6	294	1	HUPK_RHOCA
19	80.5	6.6	870	1	BCA1_HUMAN
20	80.5	6.6	1473	1	NALL_HUMAN
21	80	6.5	200	1	CNTF_RAT
22	80	6.5	422	1	Y140_HUMAN
23	80	6.5	2261	1	RRPL_MUMPM
24	79	6.4	1621	1	ALK_MOUSE
25	78.5	6.4	390	1	YL2B_STRCO
26	78.5	6.4	1102	1	CARB_STRCO
27	78	6.4	3680	1	DMD_CANFA
28	77	6.3	586	1	UL84_HCNVA
29	77	6.3	587	1	UL84_HCMVT
30	76.5	6.2	571	1	ATKA_MYCTU
31	76.5	6.2	572	1	SYM_AERPE
32	76.5	6.2	995	1	M2B2_PIG
33	76	6.2	236	1	RECO_ANASP

34	76	6.2	1001	1	PTPX_MOUSE
35	76	6.2	1132	1	BAT3_HUMAN
36	75.5	6.2	830	1	VPP3_HUMAN
37	75	6.1	199	1	CNTF_RABIT
38	75	6.1	315	1	YNEF_ECOLI
39	75	6.1	343	1	DFRA_SYNY3
40	75	6.1	450	1	TLL_DROVI
41	75	6.1	917	1	SVI_STAAU
42	75	6.1	1620	1	ALK_HUMAN
43	74.5	6.1	346	1	YG2Q_YEAST
44	74.5	6.1	409	1	NER1_MOUSE
45	74.5	6.1	444	1	CEFG_CEPAC

ALIGNMENTS

RESULT 1
CTF1_RAT
ID CTFL RAT STANDARD; PRT; 203 AA.
AC O63086;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Cardiotrophin-1 (CT-1)

GN CTF1

OS Rattus norvegicus (Rat)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Heart;

EX MEDLINE=96193659; PubMed=8604995;

RA Ishikawa M., Saito Y., Miyamoto Y., Kuwahara K., Ogawa E.,

RA Nakagawa O., Harada M., Masuda I., Nakao K.;

RT "cDNA cloning of rat cardiotrophin-1 (CT-1): augmented expression of

CT-1 gene in ventricle of genetically hypertensive rats.";

RL Biochem. Biophys. Res. Commun. 219:377-381(1996).

CC -I- FUNCTION: INDUCES CARDIAC MYOCYTE HYPERTROPHY IN VITRO. BINDS TO

AND ACTIVATES THE LEUKEMIA INHIBITORY FACTOR RECEPTOR (LIF

RECEPTOR)/GP 130 RECEPTOR COMPLEX.

CC -I- SUBCELLULAR LOCATION: Secreted (by similarity).

CC -I- TISSUE SPECIFICITY: EXPRESSED IN THE VENTRICLE AND ATRIUM OF ADULT

RATS. ALSO DETECTED IN THE LUNG, KIDNEY, LIVER, SKELETAL MUSCLE,

STOMACH AND URINARY BLADDER. NOT DETECTED IN BRAIN, COLON, TESTIS,

SPLEEN OR THYMUS. OVEREXPRESSED IN THE VENTRICLES IN THE CASE OF

HYPERTENSION AND HYPERTROPHY.

CC -I- SIMILARITY: BELONGS TO THE IL-6 SUPERFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; D78591; BAAL1427.1; -

DR InterPro; IPR001581; LIF_OSM.

DR SMART; SM00080; LIF_OSM; 1.

KW Cytochrome.

SQ SEQUENCE 203 AA; 21439 MW; DF8B921A2FALC932 CRC64;

Query Match 10.2%; Score 124.5; DB 1; Length 203;

Best Local Similarity 28.1%; Pred. No. 0.00013;

Matches 50; Conservative 24; Mismatches 85; Indels 19; Gaps 5;

QY 40 IOKYDLYRLEHQLRSLAGTYLNLGPPNEDFNPRL---GAETLPRTVDLEWRS 96

Db 27 IROTHNARLLTKYADQLLEYYQQQGEPPGLPGSPRLPLAGLSGFPSPSHAGLPV--- 83

QY 97 LNDKRLTONYEAYSHLLCYLRGINROAA-----TAEIIRSLAHFCTSLQQLLGSIGVM 151

CC similarity).

CC -!- TISSUE SPECIFICITY: Expressed in brain, testis, kidney, stomach,

CC small intestine, liver, and lung. Not detected in heart, skeletal

CC muscle, and spleen.

CC -!- MISCELLANEOUS: Mice homozygous for a null mutation of the CST gene

CC display hindlimb weakness from week 6 of age and subsequently show

CC a pronounced tremor and progressive ataxia. Myelin vacuolation is

CC observed in the cerebellar white matter, diencephalon, brainstem

CC and spinal anterior column. Male mice were infertile due to a

CC blocked spermatogenesis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB032940; BAA93009.1; -

DR EMBL; AB032939; BAA93008.1; -

DR EMBL; AK007645; BAB25160.1; -

DR EMBL; BC026806; AAB26806.1; -

DR MGD; MGI:1858277; Gcst

KW Transferase; Transmembrane; Glycoprotein.

FT DOMAIN 1 12

FT TRANSMEM 13 35

FT 1 12

FT 36 423

FT DOMAIN 36 423

FT CARBOHYD 66 66

FT CARBOHYD 312 312

FT CARBOHYD 18 18

FT CONFLICT 263 263

FT CONFLICT 271 271

FT CONFLICT 358 358

FT CONFLICT 392 392

FT CONFLICT 398 398

FT CONFLICT 423 423

SQ SEQUENCE 423 AA; 48968 MW; FD54AIA71F4AE646 CRC64;

Query Match 7.7%; Score 94; DB 1; Length 423;

Best Local Similarity 22.7%; Pred. No. 0.25;

Matches 55; Conservative 23; Mismatches 60; Indels 104; Gaps 12;

Qy 44 YDLTRYELHQRLSLAGTVINY---LGPP-----FNEP----- 72

Db 204 YDPSYNAHYLNLLFFDLGYDSSLDSPASPRVQEHILEVRRFHLVLQEVFDESILVLR 263

Qy 73 -----DNPPLGAEITLPRAT-----VDLEVWRSINDKRLTQNYEA 109

Db 264 ELLCWDLEDVLYFKLNARDSPVPLSGELYRRATAWNLLDVLRYRHFN--ASFWRKYEA 321

Qy 110 YSHLLCYLRGLNRQA-ATAELRRS---LAHFCTSLQGLLGSTIA-----GVMAA 153

Db 322 F-----GREMAREVAELRQAEHMRHCIDGGQVGAFAEQDSAMQWQPLGIKSI 373

Qy 154 LGYPLPQPLPQTEPTWTGCPAHSDFLQWDDFWLLKELQ-----TWLWRSAXDF 202

Db 374 LGYNLKSI-----GPQHEQLCRM-----LTPETQYLSDLGANLWVTKLWFLRDF 420

Qy 203 NR 204

Db 421 LR 422

RESULT 4

ID NXF1 HUMAN STANDARD; PRT; 619 AA.

AC QSUBI9; Q9UOL2; Q99799;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated protein) (mRNA export factor TAP).

GN NXF1 OR TAP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]_

RP SEQUENCE FROM N.A., AND MUTAGENESIS.

RC TISSUE=Cervical carcinoma;

RX MEDLINE=99219873; PubMed=1020158;

RA Braun I.C., Rohrbach E., Schmitt C., Izaurralde E.;

RT "TAP binds to the constitutive transport element (CTE) through a novel

RT RNA-binding motif that is sufficient to promote CTE-dependent RNA

RT export from the nucleus.";

RL EMBO J. 18:1953-1965(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99257272; PubMed=10323864;

RA Kang Y., Cullen B.R.;

RT "The human Tap protein is a nuclear mRNA export factor that contains

RT novel RNA-binding and nucleocytoplasmic transport sequences.";

RL Genes Dev. 13:1126-1139(1999).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=99384298; PubMed=10454577;

RA Bear J., Tan W., Zolotukhin A.S., Taberner C., Hudson E.A.,

RA Feiber B.K.;

RT "Identification of novel import and export signals of human TAP, the

RT protein that binds to the constitutive transport element of the type

RT D retrovirus mRNAs.";

RL Mol. Cell. Biol. 19:6306-6317(1999).

RN [4]

RP SEQUENCE FROM N.A.

RX TISSUE=Placenta;

RA Strausberg R.;

RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RA Kawakami T., Noguichi S., Itoh T., Shigeta K., Senba T., Matsumura K.,

RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,

RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,

RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,

RA Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE OF 61-619 FROM N.A.

RC TISSUE=Lymphocytes;

RX MEDLINE=97318898; PubMed=9175835;

RA Yoon D.-W., Lee H., Seol W., DeMaria M., Rosenzweig M., Jung J.U.;

RT "Tap: a novel cellular protein that interacts with tip of herpesvirus

RT saimiri and induces lymphocyte aggregation.";

RL Immunity 6:571-582(1997).

RN [7]

RP FUNCTION.

RX PubMed=9660949;

RA Grueter P., Taberner C., von Kobbe C., Schmitt C., Saavedra C.,

RA Bachi A., Wilm M., Felber B.K., Izaurralde E.;

RT "TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export

RT from the nucleus.";

RL Mol. Cell 1:649-659(1998).

RN [8]

RP CHARACTERIZATION.

RX MEDLINE=21282872; PubMed=11259411;

RA Braun I.C., Herold A., Conti E., Izaurralde E.;

RT "Overexpression of TAP/p15 heterodimers bypasses nuclear retention and

RT stimulates nuclear mRNA export.";

RL J. Biol. Chem. 276:20536-20543(2001).

RN [9]

RP CHARACTERIZATION.

RX PubMed=10668806;

RA Bachi A., Braun I.C., Rodrigues J.P., Pante N., Ribbeck K.,

RA von Kobbe C., Kutay U., Wilm M., Gorlich D., Carmo-Fonseca M.,

RA Izaurrealde E.;
 RT "The C-terminal domain of TAP interacts with the nuclear pore complex
 RL and promotes export of specific CTE-bearing RNA substrates.";
 RN RNA 6:136-158(2000).
 RP [10]
 RP MUTAGENESIS.
 RX MEDLINE=21151125; PubMed=11256625;
 RA Suyama M., Doerks T., Braun I.C., Sattler M., Izaurralde E., Bork P.;
 RT "Prediction of structural domains of TAP reveals details of its
 RL interaction with p15 and nucleoporins.";
 RN EMBO Rep. 1:53-58(2000).
 RP [11]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 102-372.
 RX MEDLINE=20514125; PubMed=11060011;
 RA Liker E., Fernandez E., Izaurralde E., Conti E.;
 RT "The structure of the mRNA export factor TAP reveals a cis arrangement
 RL of a non-canonical RNP domain and an LRR domain.";
 RN EMBO J. 19:5587-5598(2000).
 RP [12]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH NXT1, AND X-RAY
 RX CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEX WITH NXT1-FG-REPEAT.
 RA Fribourg S., Braun I.C., Izaurralde E., Conti E.;
 RT "Structural basis for the recognition of a nucleoporin FG repeat by
 RL the NTF2-like domain of the TAP/p15 mRNA nuclear export factor.";
 RN Mol. Cell 8:645-656(2001).
 RP [13]
 RP STRUCTURE BY NMR OF 551-619, AND MUTAGENESIS OF PHE-617.
 RX MEDLINE=21912422; PubMed=11875519;
 RA Grant R.P., Hurt E., Neuhaus D., Stewart M.;
 RT "Structure of the C-terminal FG-nucleoporin binding domain of
 RL Tap/NXF1.";
 RN Nat. Struct. Biol. 9:247-251(2002).
 CC -!- FUNCTION: Involved in the nuclear export of mRNA species bearing
 CC retroviral constitutive transport elements (CTE) and in the export
 CC of mRNA from the nucleus to the cytoplasm.
 CC -!- SUBUNIT: Interacts with NXT1, NXT2, E1B-AP5, RAE1, the REF
 CC proteins and with several nucleoporins.
 CC -!- SUBCELLULAR LOCATION: Nuclear; localized predominantly in the
 CC nucleoplasm and at both the nucleoplasmic and cytoplasmic faces of
 CC the nuclear pore complex. Shuttles between the nucleus and the
 CC cytoplasm.
 CC -!- TISSUE SPECIFICITY: Expressed ubiquitously.
 CC -!- DOMAIN: The minimal CTE binding domain consists of an RNP-
 CC type RNA binding domain (RBD) and leucine-rich repeats.
 CC -!- DOMAIN: The nucleoporin binding domain consists of a NTF2-like
 CC domain and a UBA-like domain. The NTF2 domain heterodimerizes with
 CC NXT1 AND NXT2. The formation of NXF1/NXT1 heterodimers is required
 CC for NXF1-mediated nuclear mRNA export. The UBA-like domain
 CC mediates direct interactions with nucleoporin-FG-repeats and is
 CC necessary and sufficient for localization of NXF1 to the nuclear
 CC rim. The conserved loop 594-NWD-596 of the UBA domain has a
 CC critical role in the interaction with nucleoporins.
 CC -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are
 CC essential for the export of mRNA from the nucleus.
 CC -!- MISCELLANEOUS: The RNA-binding domain is a non-canonical RNP-type
 CC domain.
 CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.
 CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).
 CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ132712; CAA10753.1; -.
 DR EMBL; AF12880; AAD39102.1; -.

DR EMBL; AF126246; AAD20016.1; -.
 DR EMBL; BC004904; AAO4904.1; -.
 DR EMBL; AK027192; -; NOT ANNOTATED_CDS.
 DR EMBL; U80073; AAB8111.1; -.
 DR EMBL; U80073; AAB8111.1; -.
 DR EMBL; HGNC:8071; NXF1.
 DR MIM; 602647; -.

DR InterPro; IPR001611; LRR.
 DR InterPro; IPR003603; LRRcap.
 DR PDB; 1G05; 18-OCT-01.
 DR PDB; 1JRG; 12-JUL-01.
 DR PDB; 1JNS; 23-JUL-01.
 DR PDB; 1JNS; 23-JUL-01.
 DR Pfam; PF00560; LRR; 2.
 DR PRINTS; PR00019; LEURICHRPT.
 DR SMART; SM00446; LRRcap; 1.
 DR PROSITE; PS01177; NTF2 DOMAIN; 1.
 DR Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat;
 KW Leucine-rich repeat; Multigene family; 3D-structure.
 FT DOMAIN 119 198 RNA-BINDING (RRM).
 FT REPEAT 266 291 LRR 1.
 FT REPEAT 292 315 LRR 2.
 FT REPEAT 316 343 LRR 3.
 FT REPEAT 344 371 LRR 4.
 FT DOMAIN 386 536 NTF2.
 FT DOMAIN 566 608 UBA-LIKE.
 FT DOMAIN 67 100 NUCLEAR LOCALIZATION SIGNAL.
 FT DOMAIN 83 110 NUCLEAR EXPORT SIGNAL.
 FT DOMAIN 551 561 PRO-RICH.
 FT MUTAGEN 306 308 ERE->AAA: DECREASES THE EXPORT OF MRNAS FROM THE NUCLEUS.
 FT MUTAGEN 594 594 W->A: SUPPRESSES FG-NUCLEOPORIN BINDING.
 FT MUTAGEN 595 595 D->R: SUPPRESSES FG-NUCLEOPORIN BINDING.
 FT MUTAGEN 617 617 F->A: SUPPRESSES FG-NUCLEOPORIN BINDING.
 FT CONFLICT 119 119 W -> C (IN REF. 6).
 FT CONFLICT 256 256 T -> N (IN REF. 3).
 SQ SEQUENCE 619 AA; 70182 MW; 339872AADA789FBF CRC64;

Query Match 7.6%; Score 93; DB 1; Length 619;
 Best Local Similarity 24.7%; Pred. No. 0.49;
 Matches 60; Conservative 32; Mismatches 93; Indels 58; Gaps 13;

Qy 5 AGDSGML-----ACLTVLVHLPAVPAALNRTGDP-GPGPSIQKTYDILT---RYLEHQL 54
 Db 400 SGDRGLLDVHDGACCSLSIPFQNPASSLAEYFKDSRVVKKLDPFLRLKHTR 459
 Qy 55 RSLAETLYNLGPPNEPDPNP--PRLAGETLPRATVDL-----EVRSLNDKLR-LTON 106
 Db 460 LNVV-AFLNEL--PKTQHDVNSFVVDISATSTLLCFSVNGVPEKVDGKSRDSIRAFRT 516
 Qy 107 Y---EAYSHLLCYLRG--LNROAATAELRRSLAHFCTSLQGLLGSITAGVMAALGYPLPQP 161
 Db 517 FIAVPASNSGLCIVNDELFFVRNASSEIQRAFA-----MPAP 553
 Qy 162 LPTGPTTPGPAHSDFLQK-----MDDFWLLKELQTLW---RSKADFNRLKKMQPP 212
 Db 554 TFSSEFVPTLSPQQEMLQAFSTQSGNLEWSQKCLQDNNDYTRSAQAATHLKAKGEIP 613
 Qy 213 AAA 215
 Db 614 EVA 616

RESULT 5
 CTF1 HUMAN STANDARD; PRT; 201 AA.
 ID CTF1 HUMAN STANDARD; PRT; 201 AA.
 AC Q16619;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cardiotrophin-1 (CT-1).
 GN CTF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92304573; PubMed=1610564;
 RA Leung D.W., Parent A.S., Cachianes G., Lee A.L., Nikolics K.,
 RA Esch F., Coulombe J.N., Blacher R.W., Eckenstein F.P., Nishi R.;
 RT "Cloning, expression during development, and evidence for release of
 RT a trophic factor for ciliary ganglion neurons";
 RL Neuron 8:1045-1053(1992).
 CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
 CC AXOTOMY.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80827; AAA48784.1; -;
 DR PIR; JH0680; JH0680.
 DR HSSP; F26441; ICNT.
 DR InterPro; IPR000151; Ctl_neuro_factor.
 DR Pfam; PF01110; CNTF; 1.
 DR ProDom; PD011041; Ctl_neuro_factor; 1.
 DR Growth factor; Neurone.
 KW Growth factor; Neurone.
 SQ SEQUENCE 195 AA; 21330 MW; FEA076949DB34AC5 CRC64;
 Query Match 7.4%; Score 91; DB 1; Length 195;
 Best Local Similarity 27.3%; Pred. No. 0.18;
 Matches 51; Conservative 21; Mismatches 83; Indels 32; Gaps 9;
 QY 46 LRYLEHQLRSAGTYLVYLPFPNPPRLGAEPLPRATVLEWVRSINDKRLTLQ 105
 DB 23 LARKMSDVTDLDDIYVERQG-----LDASISVAADVGVTAIV--ERWAEQTGTQLLD 75
 QY 106 N---YEAYSHLLCYLRGLNRQA---ATAELRSLA-----HFCTSLOGLLGSIAGWMA 152
 DB 76 NLAAVFAFRTLQAQMLEEQRELLGDDTAELGPALAAWLLQSAFVYHEELL-----ELE 130
 QY 153 ALGYPLPQPLPGTEPTWTPGPAH-SPLQKMDDFWLKELQTLWRSKADFNRLKKQNP 211
 DB 131 SRGAPAE---GSEP---PAPPLSLFQKLRGLRVLRLAQWAVRSVRDLRLSLKHGPG 184
 QY 212 PAAAVTL 218
 DB 185 SGAALGL 191
 RESULT 8
 MIP1_SCHPO STANDARD; PRT; 1313 AA.
 AC P87141;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE WD-repeat protein mipl.
 GN MIP1 OR SPAC57A7.11.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.; FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20115869; PubMed=10648609;
 RA Shinozaki-Yabana S., Watanabe Y., Yamamoto M.;

RT "Novel WD-repeat protein Miplp facilitates function of the meiotic
 RT regulator Mei2p in fission yeast.";
 RL Mol. Cell. Biol. 20:1234-1242(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., William R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: BINDS TO AND FACILITATES THE FUNCTIONING OF THE MEIOTIC
 CC REGULATOR MEI2. MAY ALSO BE INVOLVED IN CONJUGATION BY INTERACTING
 CC WITH STELL. ESSENTIAL FOR CELL GROWTH.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB032552; BAA84595.1; -;
 DR EMBL; Z95396; CAB08769.1; -;
 DR InterPro; IPR001680; WD40.
 DR InterPro; IPR004083; Yeast176.
 DR Pfam; PF00400; WD40; 5.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR PRINTS; PR01547; YEAST176DUF.
 DR SMART; SM00320; WD40; 5.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS00682; WD_REPEATS_2; 2.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Meiosis; WD repeat; Repeat.
 FT REPEAT 986 1029 WD 1.
 FT REPEAT 1033 1074 WD 2.
 FT REPEAT 1087 1126 WD 3.
 FT REPEAT 1130 1170 WD 4.
 FT REPEAT 1176 1216 WD 5.
 FT REPEAT 1219 1259 WD 6.
 FT REPEAT 1268 1308 WD 7.
 SQ SEQUENCE 1313 AA; 148533 MW; C71B663B0171E7A4 CRC64;
 Query Match 7.3%; Score 89.5; DB 1; Length 1313;
 Best Local Similarity 22.8%; Pred. No. 2.7;
 Matches 61; Conservative 32; Mismatches 69; Indels 105; Gaps 14;
 QY 15 LCTVLWHLPAV--PALNR-----TGDPGPGPSI-----QKTYDLTRYLEHQLRS 56


```
Db 811 LAFLLQHLPALHAKASLKSDTNSVTSDPKPHFPVPSVSENKILNRSFSLRSLKGLALS 870
QY 57 LAG-----TVLNYL-----GPP-FNEPDFNPPRLGAETLPRATVDL 91
Db 871 LAGSDRASELLSNGENKPAESLNHLTSAKVGPFPAPFNELEY-----QSELDW 919
QY 92 EVWRSND-----KRLTQNYEAYSHLLCYLRGL-----NRQAATAELRRSLAHFCTS 139
Db 920 PLTSYLFDSRSKYFTEPQMRPNEDDEPGS--ICYNQLWRNENKLIYRTRPLAEYSTN 977
QY 140 -----LQGLLSGIAGVMAALGVPLOPLEGTEPTWPGPAHSDFLQKMDDFWLKE--- 190
Db 978 GRWNOQLMTFNNTIA-----PRKLMFHQFEDQLITLGDKDI 1013
QY 191 LQWLWRSKDNRLKKMKOPPPAAAVT 217
Db 1014 IQVDWR-----NRLNSFKTSASAT 1036
RESULT 9
HAIR_RAT
ID HAIR_RAT STANDARD; PRT; 1181 AA.
AC F97609;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hairless protein.
GN HR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=97141510; PubMed=8987811;
RA Thompson C.C.;
RT "Thyroid hormone-responsive genes in developing cerebellum include a
RL novel synaptotagmin and a hairless homolog.";
RL J. Neurosci. 16:7832-7840(1996).
CC -!- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO
CC REGULATE ONE OF THE PHASES OF HAIR GROWTH.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U71293; AAC53018.1; ALT INIT.
DR InterPro; IPR003347; TF_JmJC.
DR Pfam; PF02373; jmjC; 1.
KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;
KW Metal-binding.
FT ZN FING 594 619 C6-TYPE.
SQ SEQUENCE 1181 AA; 127307 MW; 834B7029CF8E88F0 CRC64;
Query Match 7.1%; Score 87.5; DB 1; Length 1181;
Best Local Similarity 25.3%; Pred. No. 3.6;
Matches 61; Conservative 21; Mismatches 92; Indels 67; Gaps 12;
QY 30 RTGDPGSGPSIQKTYDITRYLEHQLRSIAGTY--LNVIGPP-----FNEPDF 74
Db 859 RQGPVLVSGTQKTLRLSLMGWELGQVQLTALGPPQPTSLDSTAFKGFSPHEA 918
QY 75 NPPELGAETLPRATVDLEVRSLNDK-----LRLTQ-----NREAYSHLIC 115
Db 919 RP-----KLDEGSV-LLLHRPLGDKDESRENVENLASSLPUEYCAHQKUNLASLPLGL 971
```

```
QY 116 YLRGLNRQAATAELRRS-----LAHFTSLOGLLSIAGVMAALGYPLPQPLGTEPTW 169
Db 972 TLHLEPQLWAAVGVNHRGHLGTLKNCVSESDLSILVHAEAL-----PPW 1019
QY 170 TPGPAHSDFLQKMD--DFWLLKELQTLWR--SAKDFNRLKKMKOP--PAAAVTLHLGAH 223
Db 1020 Y--RAQKDFLSGLDGEGLWSPGSQTSTVHVHVFRAQDAQRIRRELQWVCPAGAGTLERGAP 1077
QY 224 G 224
Db 1078 G 1078
RESULT 10
M2B2 HUMAN
ID M2B2 HUMAN STANDARD; PRT; 1009 AA.
AC Q9Y2E5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epididymis-specific alpha-mannosidase precursor (EC 3.2.1.24)
DE (Mannosidase alpha class 2B member 2).
GN MAN2B2 OR KIAA0935.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Stone N.E., Schmutz J.J., Cox D.R., Myers R.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 132-1009 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [3]
RP RECONSTRUCTION FROM GENOMIC SEQUENCE.
RA Bairoch A.;
RL Unpublished observations (NOV-2001).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC mannose residues in alpha-D-mannosides.
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC004480; -- NOT ANNOTATED CDS.
DR EMBL; AB023152; BAA76779.1; ALT_SEQ.
KW Hydrolase; Glycosidase; Signal; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1009 EPIDIDYMIS-SPECIFIC ALPHA-MANNOSIDASE.
FT CARBOHYD 226 226 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 808 808 N-LINKED (GLCNAC. .) (POTENTIAL).
```


FT CARBOHYD 812 812 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1009 AA; 113987 MW; F3DB81DD061352E6 CRC64;
Query Match 7.0%; Score 86; DB 1; Length 1009;
Best Local Similarity 23.3%; Pred. No. 4.1;
Matches 45; Conservative 20; Mismatches 66; Indels 62; Gaps 8;
QY 50 LEHQ-----LRSAGTLYNLGPPNPPRPRGAEITLPRATVLEWVRSNDKRLTQ 105
DB 842 LQHRPVVGLDLAGTAPKLPFGPOQQAVALTPPNLHLQIL-----SIPGWRYSNHTSHSQ 896
QY 106 NYEAYSHLLCYLRGLNRQAATAELRRSLAHF-----CTSLQGLLGSIAQ 149
DB 897 N-----LRKGRGEAQADLRRLVRLYHLXYEVGEDPVLSPQVTVNLQAVLQALGS 946
QY 150 VMAALGYPLPOPLPGCTEPTWPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNR--LKK 207
DB 947 VVAV-----BERSLTGT-----WDLSSLHRWSWRTGGRHRGDTTS 982
QY 208 KMOPPA--AVTLH 219
DB 983 PSRPPGPIITVH 995
RESULT 11
PHAC PSEOL STANDARD; PRT; 560 AA.
AC P26456;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Poly(3-hydroxyalkanoate) polymerase 2 (EC 2.3.1.-) (PHA polymerase 2)
DE (PHA synthase 2) (Polyhydroxyalkanoic acid synthase 2).
GN PHAC.
OS Pseudomonas oleovorans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI TaxID=301;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GP01;
RX MEDLINE=91115830; PubMed=1989978;
RA Huismen G.W., Wonik E., Meima R., Kazemier B., Terpstra P.,
RA "Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas
RT oleovorans. Identification and sequences of genes and function of the
RT encoded proteins in the synthesis and degradation of PHA."
RL J. Biol. Chem. 266:2191-2198(1991).
CC -!- FUNCTION: P.OLEOVRANS ACCUMULATES POLY(3-HYDROXYALKANOATES) AFTER
CC GROWTH ON MEDIUM CHAIN LENGTH HYDROCARBONS. LARGE AMOUNTS OF THIS
CC POLYESTER ARE SYNTHESIZED WHEN CELLS ARE GROWN UNDER NITROGEN-
CC LIMITING CONDITIONS. WHEN NITROGEN IS RESUPPLIED IN THE MEDIUM,
CC THE ACCUMULATED PHA IS DEGRADED.
CC -!- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M58445; AAA25934.1; --
DR PIR; C38604; C38604.
DR InterPro; IPR000073; Abhydrolase.
DR Pfam; PF00561; abhydrolase; 1.
KW PHA biosynthesis; Transferase; Acyltransferase.
FT ACT SITE 296 296 POTENTIAL.
SQ SEQUENCE 560 AA; 62631 MW; E2CD844FC1616B83 CRC64;
Query Match 7.0%; Score 85.5; DB 1; Length 560;

Best Local Similarity 30.2%; Pred. No. 2.2;
Matches 38; Conservative 17; Mismatches 42; Indels 29; Gaps 7;
QY 77 PRLGAETLPRATVLEWVRSNDKRLTQNYEAYSHLLCYLRGLNRQAATAELRRSL--A 134
DB 5 PAKGTPTLPATSMNVQ-----NAILGLRGR-----DLISLTLNRVRSQS-----LRHPLHTA 50
QY 135 HFCTSLQGLLGSIAQVMAALGYPLPQPLFG-----TEPTWTPGPAHSDFLQKMDDFWLLKE 190
DB 51 HHLIALGSGQLGRV-----ILGDTPLQPNPRDPRPSDPTWSQNPFFYRRGLQA-----YLAQW 101
QY 191 LQTLW 196
DB 102 KQTRLW 107
RESULT 12
CNTF_PIG STANDARD; PRT; 200 AA.
AC O02732;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ciliary neurotrophic factor (CNTF).
GN CNTF.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98030048; PubMed=9363597;
RA Weakes R.L., Ramsoondar J.J., Gallagher D.S. Jr., Noguees C.,
RA Piedrahita J.A.;
RT "Isolation, characterization and chromosomal localization of the
RT porcine ciliary neurotrophic factor (CNTF) gene."
RL Anim. Genet. 28:354-357(1997).
CC -!- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
CC AXOTOMY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: NERVOUS SYSTEM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CNTF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U57644; AAC27342.1; --
DR HSP; P26441; ICNT.
DR InterPro; IPR000151; Ctl_neuro_factor.
DR Pfam; PF01110; CNTF; 1.
DR ProDom; PD011041; Ctl_neuro_factor; 1.
KW Growth factor; Neurone.
SQ SEQUENCE 200 AA; 22718 MW; 40507C4457ED6531 CRC64;
Query Match 6.9%; Score 84; DB 1; Length 200;
Best Local Similarity 24.7%; Pred. No. 0.84;
Matches 47; Conservative 23; Mismatches 88; Indels 32; Gaps 9;
QY 46 LTRYLEHQRLSLAGTLYNLGPPNPPRPRGAEITLPRATVLEWVRSNDKRLTQ 105
DB 23 LARKIRSDLTALMEAYVKHQG--LNE---NINLDSVDGVPMASDR--WSELTAEERLQ 75
QY 106 NYEAYSHLLCYLRGLNRQAATAELRRSLAHFCTS---LQGLLGSIAQVMAALGYPLPQ-- 160
DB 76 NLRAYRTFHVMLARL-----LEDQREHFTPAEDDDHQAHTIVLVQAAFAAYQLEELM 127
QY 161 -----PLPGTEFTWTPGPAHSDF--FLQKMDDFWLLKELQTLWRSKDFNRLLKKMQPPA 213

Db 128 VLEHKKVPPSADGTPLSVGGGLFEKKLWGLKVLQELSQWTVSRDLRVISHQ----- 183
 QY 214 AAVTLHLGAH 223
 Db 184 AGVPAH-GSH 192

RESULT 13
 TLL_DROME
 ID TLL_DROME STANDARD; PRT; 452 AA.
 AC P18102; Q9VA33;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Tailless protein.
 GN TLL OR NR2E2 OR CG1378.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=90304905; PubMed=2364433;
 RA Pignoni F., Baldarelli R.M., Steingrimsson E., Diaz R.J.,
 RA Patapoutian A., Merriam J.R., Lengyel J.A.;
 RT "the Drosophila gene tailless is expressed at the embryonic termini
 RT and is a member of the steroid receptor superfamily.";
 RL Cell 62:151-163 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93157371; PubMed=8430097;
 RA Liaw G.-J., Steingrimsson E., Pignoni F., Courey A.J., Lengyel J.A.;
 RT "Characterization of downstream elements in a Raf-1 pathway.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:858-862 (1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Minkov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskaas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=99287803; PubMed=10357938;
 RA Daniel A., Dumstrei K., Lengyel J.A., Hartenstein V.;
 RT "The control of cell fate in the embryonic visual system by atonal,
 RT tailless and EGFR signaling.";
 RL Development 126:2945-2954 (1999).
 CC -|- FUNCTION: ORPHAN RECEPTOR THAT BINDS DNA AS A MONOMER TO HORMONE
 CC RESPONSE ELEMENTS (HRE) CONTAINING AN EXTENDED CORE MOTIF HALF-
 CC SITE SEQUENCE 5'-AAGTCA-3' IN WHICH THE 5' FLANKING NUCLEOTIDES
 CC PARTICIPATE IN DETERMINING RECEPTOR SPECIFICITY. THIS RECEPTOR
 CC BINDS TO THE CONSENSUS SEQUENCE [AG][AG]AAGTCAA. PLAYS A KEY ROLE
 CC IN THE ESTABLISHMENT OF NONMETAMERIC DOMAINS AT THE ANTERIOR AND
 CC POSTERIOR POLES OF THE EMBRYO. IT MAY ALSO PLAY A ROLE IN THE
 CC NERVOUS SYSTEM. THE MATERNAL TERMINAL PATHWAY ACTIVATES THE TLL
 CC GENE IN THE TERMINI; TLL ACTIVITY THEN REPRESSES SEGMENTATION AND
 CC ACTIVATES TERMINAL-SPECIFIC GENES IN THESE DOMAINS. INVOLVED IN
 CC THE REGULATION OF EARLY EYE DEVELOPMENT. IN THE EMBRYONIC VISUAL
 CC SYSTEM ANLAGE DRIVES CELLS TO OPTIC LOBE AS OPPOSED TO BOLWIG'S
 CC ORGAN FATE.
 CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -|- TISSUE SPECIFICITY: BRAIN AND PERIPHERAL NERVOUS SYSTEM.
 CC -|- DEVELOPMENTAL STAGE: DURING STAGE 10 FOUND IN THE ANTERIOR PART OF
 CC THE VISUAL SYSTEM THAT LATER GIVES RISE TO THE ANTERIOR LIP OF THE
 CC OPTIC LOBE. AT STAGE 12 ALSO FOUND IN THE POSTERIOR LIP OF THE
 CC OPTIC LOBE. IN THIRD LARVAL INSTAR EXPRESSED IN THE OPTIC LOBE OF
 CC THE LARVAL BRAIN AND IN THE EYE ANTENNAL DISK, BOTH IN ANTENNAL
 CC AND EYE PORTION.
 CC -|- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
 CC NR2 SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@sib-sib.ch.
 CC -----

EMBL; M34639; AAA28936.1; --
 EMBL; AF019362; AAB71371.1; --
 EMBL; AE003775; AAF57091.1; --
 PIR; A35602; A35602.
 PIR; A47265; A47265.
 HSP; P10826; JHRA.
 TRANSFAC; T00789; --
 FlyBase; FBgn003720; tll.
 InterPro; IPR000536; Hormone_rec_lig.
 InterPro; IPR001628; Znf_C4steroid.
 Pfam; PF00104; hormone_rec; 1.
 Pfam; PF00105; zf-C4; 1.
 PRINTS; PR00047; STROIDFINGER.
 ProDom; PD000035; Znf_C4steroid; 1.
 SMART; SM00430; HOL1; 1.
 SMART; SM00399; Znf_C4; 1.
 PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 Zinc-finger; Activator; Repressor; Developmental protein.
 DNA_BIND 34 101 NUCLEAR RECEPTOR-TYPE.
 ZN_FING 34 54 C4-TYPE.
 ZN_FING 70 96 C4-TYPE.
 DOMAIN 244 389 LIGAND-BINDING (BY SIMILARITY).
 DOMAIN 262 265 POLY-LEU.
 SQ SEQUENCE 452 AA; 50549 MW; A4ABEFFDE993A37C CRC64;

Query Match 6.8%; Score 83; DB 1; Length 452;

[illegible]

RESULT 14

RES001	HAIR_HUMAN	HAIR_HUMAN	STANDARD;	PRT; 1189 AA.
ID	AD	O43593;		
DT	15-JUL-1999	(Rel. 38, *Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE		Hairless protein.		
GN	HR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI	NCBI_TaxID=9606;			
RP	[1]	SEQUENCE FROM N.A. (LONG ISOFORM), AND VARIANT ALUNC ALA-1022.		
RP	MDLNE=98111413; PubMed=9445480;			
RX	Ahmad W., Ul Haque M.F., Brancolini V., Tsou H.C., Ul Haque S.,			
RA	Lam H., Alta V.M., Owen J., Dblaquiere M., Frank J.,			
RA	Cserhalmai-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,			
RA	Ahmad M., Ott J., Christiano A.M.;			
RT	"Alpecia universalis associated with a mutation in the human hairless			
RT	gene.";			
RL	Science 279:720-724(1998).			
RP	[2]			
RP	SEQUENCE FROM N.A. (LONG ISOFORM), REVISIONS TO 572 AND 774, AND			
RP	TISSUE SPECIFICITY.			
RC	TISSUE=Peripheral blood leukocytes, and Skin fibroblast;			
RC	MDLNE=99162400; PubMed=10051399;			
RX	Ahmad W., Zlotogorski A., Panteleyev A.A., Lam H., Ahmad M.,			
RA	ul Haque M.F., Abdallah H.M., Dregan L., Christiano A.M.;			
RA	"Genomic organization of the human hairless gene (HR) and			
RT	identification of a mutation underlying congenital atrichia in an Arab			
RT	Palestinian family.";			
RL	Genomics 56:141-148(1999).			
RP	[3]			
RP	SEQUENCE FROM N.A., VARIANT ALUNC ASP-1136, TISSUE SPECIFICITY, AND			
RP	ALTERNATIVE SPLICING.			
RC	TISSUE=Peripheral blood leukocytes, Brain, and Fetal brain;			
RC	MDLNE=9840496; PubMed=9736769;			
RX	Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillmer A.,			
RA	Farooq S.A., Al-Dhafri K.S., Ahmad M., Haque S., Rietschel M.,			
RA	Propping P., Kruse R., Noethen M.M.;			
RT	"Cloning, genomic organization, alternative transcripts and mutational			
RT	analysis of the gene responsible for autosomal recessive universal			
RT	congenital alopecia.";			
RL	Hum. Mol. Genet. 7:1671-1679(1998).			
RP	[4]			
RP	VARIANT APL GLN-620.			
RP	MDLNE=98431781; PubMed=9758627;			
RX	Ahmad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,			
RA	Panteleyev A.A., Ahmad M., McGrath J.A., Christiano A.M.;			
RA	"A missense mutation in the zinc-finger domain of the human hairless			
RT	gene underlies congenital atrichia in a family of Irish travellers.";			
RT	Am. J. Hum. Genet. 63:984-991(1998).			
CC	-1- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO			
CC	REGULATE ONE OF THE PHASES OF HAIR GROWTH.			
CC	-1- SUBCELLULAR LOCATION: Nuclear.			

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:17 ; Search time 33 Seconds
(without alignments)
908.527 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGLACTLTVL.....KKMQPPAAAVTLHGAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq 101002.*

1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1226	100.0	225	19 AAW29715	Human neurotrophic
2	1226	100.0	225	19 AAW56141	Amino acid sequenc
3	1226	100.0	225	20 AAW94466	Human cardiotothro
4	1226	100.0	225	21 AAY87813	Human NNT-1 protei
5	1226	100.0	225	22 AAG63543	Amino acid sequenc
6	1226	100.0	225	23 AAY78176	Human novel neurot
7	1226	100.0	225	22 AAM25831	Human protein sequ
8	1214	99.0	223	22 AAE00828	Human cardiotothro
9	1204	98.2	321	22 ABB11896	Human cardiotothro
10	1204	98.2	321	22 AAM79399	Human protein SEQ

11	1197	97.6	260	22 AAM78415	Human protein SEQ
12	1193	97.3	225	19 AAW29716	Mouse neurotrophic
13	1193	97.3	225	19 AAW56142	Amino acid sequenc
14	1193	97.3	225	21 AAY87814	Murine NNT-1 prote
15	1193	97.3	225	23 AAY78177	Mouse novel neurot
16	1169	95.4	215	21 AAB19586	Human interleukin-
17	1136	92.7	215	21 AAB19587	Mouse interleukin-
18	885	72.2	164	22 ABB40317	Peptide #7823 enco
19	885	72.2	164	22 ABB24716	Protein #6715 enco
20	885	72.2	164	22 AAM61118	Human brain expres
21	885	72.2	164	22 AAM73827	Human bone marrow
22	885	72.2	164	22 AAM20115	Peptide #6549 enco
23	885	72.2	164	22 AAM34012	Peptide #8049 enco
24	885	72.2	164	23 AAG43716	Human peptide enco
25	162.5	13.3	208	20 AAY09197	Human DNAX interle
26	160	13.1	208	20 AAY09196	Human DNAX interle
27	118.5	9.7	203	16 AAR83965	Mouse cardiac hype
28	118.5	9.7	203	17 AAR88204	Human cardiotothro
29	118.5	9.7	203	18 AAR29237	Murine cardiotothro
30	96.5	7.9	243	22 AAU09153	Human cytokine Zal
31	96.5	7.9	243	22 AAB20277	Human interleukin
32	96.5	7.9	243	23 AAU76375	Human helical prot
33	92	7.5	332	21 AAG22132	Arabidopsis thalia
34	92	7.5	332	21 AAG40321	Arabidopsis thalia
35	91.5	7.5	201	16 AAR83967	Human cardiac hype
36	91.5	7.5	201	18 AAY29238	Human cardiotothro
37	91.5	7.5	201	20 AAY06490	Human tumour-associ
38	91.5	7.5	201	21 AAB27662	Human protein PR08
39	91.5	7.5	201	21 AAB33004	Human cardiotothro
40	91.5	7.5	201	21 AAY93697	Amino acid sequenc
41	91.5	7.5	201	21 AAY87818	Human cardiotothro
42	91.5	7.5	201	22 AAB50994	Human PR0882 prote
43	91.5	7.5	1182	23 AAE19798	Mouse Hairless pro
44	91	7.4	195	14 AAR34432	Sequence of growth
45	91	7.4	195	20 AAM83337	Chicken ciliary ne

ALIGNMENTS

RESULT 1
AAW29715
ID AAW29715 standard; Protein; 225 AA.
XX
AC AAW29715;
XX
DT 09-NOV-1998 (first entry)
XX
DE Human neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW peripheral neuropathy; dystrophy; neural retina degeneration;
KW common variable immunodeficiency; CVID; selective IGA deficiency;
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
XX therapy.

OS Homo sapiens.

XX
FH Key
FT Peptide
FT Protein
FT Location/Qualifiers
1..27
/label= sig_peptide
28..225
/label= Mat_protein

XX WO9833922-A1.

XX 06-AUG-1998.

XX 02-FEB-1998; 98WO-US02363.

XX 30-JAN-1998; 98US-0016534.

```

PR 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
PA
XX Chang M, Elliot GS, Sarmiento U, Senaldi G;
XX N-PSDB; AAV47510-11.
XX WPI; 1998-437475/37.
XX DR N-PSDB; AAV47510-11.
XX
XX Newly isolated nucleic acid encoding human or murine neurotrophic
PT factor NNT-1 - useful for treatment of neurological and
PT immunological diseases or inflammation, also as vaccine adjuvant
XX
XX Claim 12; Fig 3; 120pp; English.
XX
XX This is the amino acid sequence of a novel neurotrophic factor,
CC designated NNT-1, that is a growth factor for neurons and for B or
CC T cells. It was deduced from isolated cDNA (see AAV47510) and
CC genomic DNA (see AAV47511) clones. Vectors containing the cDNA or
CC genomic DNA and host cells are provided for use in the production
CC of NNT-1 polypeptides. These are used to treat: (i) neurological
CC or immunological diseases, specifically Alzheimer's, Parkinson's
CC or Huntington's diseases, amyotrophic lateral sclerosis,
CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
CC degeneration of the neural retina, or conditions characterised by T
CC or B cell defects, e.g. common variable immunodeficiency (CVID),
CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
CC inflammation. NNT-1 is also able to boost immunoreactivity and
CC antibody production following vaccination, and, since it inhibits
CC tumour necrosis factor production, it may also be useful for
CC treating sepsis. In addition, cells that have been engineered to
CC express NNT-1 can be implanted, or nucleic acids are delivered in
CC gene therapy vectors.
XX
XX SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 19; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
QY 61 YLNYLGPPNEPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPNEPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
QY 181 KMDDFWLLKELOTWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225

RESULT 2
AAW56141
ID AAW56141 standard; Protein; 225 AA.
XX AC AAW56141;
XX DT 13-JUL-1998 (first entry)
XX DE Amino acid sequence of human neurotrophic factor NNT-1.
XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
XX Homo sapiens.
XX

```

```

XX Key Location/Qualifiers
FH Peptide 1..27
FT Protein /note= "signal peptide"
FT Protein 28..225 /note= "mature protein"
XX
XX US5741772-A.
XX
XX 21-APR-1998.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
XX Chang M;
XX WPI; 1998-260526/23.
XX N-PSDB; AAV22652.
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
PT useful for stimulating growth of motor and sympathetic neurons
XX
XX Claim 1; Fig 3; 41pp; English.
XX The present sequence represents a human neurotrophic factor, designated
CC NNT-1, which is capable of stimulating growth of motor or sympathetic
CC neurons. The NNT-1 protein is useful in the treatment of neurological
CC diseases characterised by the degeneration and death of particular
CC classes of neurons. These diseases specifically include Parkinson's
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
CC stroke and various degenerative disorders affecting vision.
XX
XX SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 19; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLEHQLRSLAGT 60
QY 61 YLNYLGPPNEPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPNEPDNPPRLGAETLPRAVDLEVWRSNDKRLTONYEAYSHLLCYLRGL 120
QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
QY 181 KMDDFWLLKELOTWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225

RESULT 3
AAW94466
ID AAW94466 standard; Protein; 225 AA.
XX AC AAW94466;
XX DT 22-APR-1999 (first entry)
XX DE Human cardiostrophin-like cytokine protein.
XX Human; cardiostrophin-like cytokine; interleukin 6 cytokine family;
KW CLC; IL-6; diagnosis; detection; immune system-related disorder;
KW cancer; cardiac disorder; heart failure; hypertension; cancer;
KW autoimmune disorder; infection.
XX

```



```

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein 28..225
FT Domain 74..79
FT /label= CD-I
FT /note= "conserved domain"
FT Domain 150..156
FT /label= CD-II
FT /note= "conserved domain"
FT Domain 194..198
FT /label= CD-III
FT /note= "conserved domain"
XX
XX WO9900415-A1.
XX
XX 07-JAN-1999.
XX
XX 29-JUN-1998; 98WO-US13129.
XX
XX 30-JUN-1997; 97US-0051311.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Shi Y;
XX
XX WPI; 1999-095678/08.
XX
XX N-PSDB; AAX16161.
XX
XX New isolated cardiotrophin-like cytokine nucleic acid - used to
XX develop products for treating cardiac and immune system disorders,
XX e.g. heart failure, hypertension, cancers, autoimmune disorders and
XX infections
XX
XX Claim 1; Fig 1; 103pp; English.
XX
XX The present invention relates to a novel cardiotrophin-like cytokine
XX (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine
XX family. The present sequence represents the human CLC protein. The
XX present invention also describes screening methods for identifying
XX agonists and antagonists of CLC activity, as well as methods for
XX detecting cardiac and immune system-related disorders and
XX therapeutic methods for treating cardiac and immune system-related
XX disorders, e.g. heart failure, hypertension, cancers, autoimmune
XX disorders and infections.
XX
XX Sequence 225 AA;
XX
XX Query Match 100.0%; Score 1226; DB 20; Length 225;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-118;
XX Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 MDLRAGDSWGMACLTCTVWHLPAVPAALNRTGDPGPGSIQKTYDLYLEHQLSLAGT 60
XX Db 1 MDLRAGDSWGMACLTCTVWHLPAVPAALNRTGDPGPGSIQKTYDLYLEHQLSLAGT 60
XX
XX Qy 61 YLNYLGPPFPNEPDPNPRGLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
XX Db 61 YLNYLGPPFPNEPDPNPRGLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
XX
XX Qy 121 NQOATAELRSIAHFCCTSLQGLGSIAGVMAALGYPIPLPGTEPTWTGPAHSDFLQ 180
XX Db 121 NQOATAELRSIAHFCCTSLQGLGSIAGVMAALGYPIPLPGTEPTWTGPAHSDFLQ 180
XX
XX Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAAATVHLGHAGF 225
XX Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAAATVHLGHAGF 225
XX
XX RESULT 4

```

```

AA87813
ID AAY87813 standard; Protein; 225 AA.
XX
AC AAY87813;
XX
DT 24-AUG-2000 (first entry)
XX
DE Human NNT-1 protein.
XX
XX NNT-1; human; neurotrophic factor; neurotropic; neuroprotective; treatment;
XX anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
XX nervous system degeneration; Alzheimer's disease; Parkinson's disease;
XX amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
XX Huntington's disease; peripheral neuropathy; neural retina degeneration;
XX retinopathy; immune disorder; hematopoietic disorder.
XX
OS Homo sapiens.
XX
XX US6054294-A.
XX
XX 25-APR-2000.
XX
XX 12-DEC-1997; 97US-0988819.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX (AMGE-) AMGEN INC.
XX
XX Chang M;
XX
XX WPI; 2000-338492/29.
XX
XX N-PSDB; AAA39481.
XX
XX New nucleic acids encoding neurotrophic factors useful for stimulating
XX growth of motor or sympathetic neurons for treating neuron cell damage
XX
XX Claim 1c; Fig 3; 42pp; English.
XX
XX This invention describes a novel nucleic acid molecule (I) encoding a
XX novel neurotrophic factor (NNT-1) (II) which has neurotropic,
XX neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
XX ophthalmological activity. (I) is useful for producing NNT-1
XX polypeptides which are useful for treating patients in whom various
XX cells of the central, autonomic, or peripheral nervous system have
XX degenerated and/or have been damaged by congenital disease, trauma,
XX mechanical damage, surgery, stroke, ischemia, infection, metabolic
XX disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
XX proteins are used to treat diseases like Alzheimer's, Parkinson's,
XX amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
XX disease, peripheral neuropathy induced by diabetes or other metabolic
XX disorders, and/or dystrophies or degeneration of the neural retina such
XX as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
XX night blindness, progressive cone-rod degeneration, immune disorders and
XX hematopoietic disorders. (I) is effective in treating neurological
XX conditions and promotes neuron regeneration. Neural functions are
XX effectively restored in patients suffering from various neurological
XX disorders. This sequence represents the human NNT-1 protein described in
XX the method of the invention.
XX
XX Sequence 225 AA;
XX
XX Query Match 100.0%; Score 1226; DB 21; Length 225;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-118;
XX Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 MDLRAGDSWGMACLTCTVWHLPAVPAALNRTGDPGPGSIQKTYDLYLEHQLSLAGT 60
XX Db 1 MDLRAGDSWGMACLTCTVWHLPAVPAALNRTGDPGPGSIQKTYDLYLEHQLSLAGT 60
XX
XX Qy 61 YLNYLGPPFPNEPDPNPRGLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
XX Db 61 YLNYLGPPFPNEPDPNPRGLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120

```

CC assist blastocyst implantation, thrombosis, or retinal disease,
CC particular retinal pigmentosis.
XX
SQ Sequence 225 AA;
Query Match 100.0%; Score 1226; DB 22; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDLRAGDSWGMCLACTVILWHLPAVPAALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60
DB 1 MDLRAGDSWGMCLACTVILWHLPAVPAALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60
QY 61 YLNYLGPPFNEPDPNPRLGAEATLPRAVTVLEWMSLNDKRLTQNYEAYSHLLCYLRGL 120
DB 61 YLNYLGPPFNEPDPNPRLGAEATLPRAVTVLEWMSLNDKRLTQNYEAYSHLLCYLRGL 120
QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180
DB 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225
DB 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225
RESULT 6
AAU78176 standard; Protein; 225 AA.
ID AAU78176
XX
AC AAU78176;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human novel neurotrophic factor NNT1.
XX
KW Human; NNT1; neurotrophic factor; IgE-related disease;
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
KW vascular stenosis; rheumatoid arthritis; psoriatic arthritis;
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
KW inflammatory bowel disease; transplant rejection; reproductive disorder;
KW graft versus host disease; infertility; miscarriage; preterm labour.
XX
OS Homo sapiens.
XX
PN WO200215977-A2.
XX
PD 28-FEB-2002.
XX
PF 17-AUG-2001; 2001WO-US25906.
XX
PR 18-AUG-2000; 2000US-226436P.
PR 16-AUG-2001; 2001US-0931704.
XX
PA (AMGE-) AMGEN INC.
XX
PI Senaldi G;
XX
XX WPI; 2002-280867/32.
DR N-PSDB; ABK11647.
XX
XX Treating Immunoglobulin E-related disease, modulating IgE levels in a
PT patient; preventing IgE-related disease and treating allergic diseases,
PT involves administering NNT-1 inhibitor to a patient -
XX
XX Claim 2; Fig 3; 63pp; English.
XX
XX The invention relates to treating Immunoglobulin E (IgE)-related disease,
CC modulating IgE levels in a patient, preventing an IgE-related disease,
CC and treating allergic diseases, comprising administering a
CC therapeutically effective amount of novel neurotrophic factor (NNT)-1

QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180
DB 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTPGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225
DB 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225
RESULT 5
AAG63543 standard; Protein; 225 AA.
XX
AC AAG63543;
XX
DT 15-OCT-2001 (first entry)
XX
DE Amino acid sequence of a human NNT-1 protein.
XX
KW NNT-1; CLF-1; sCNTFRalpha; nervous system; neuron; nervous system;
KW neuro-muscular function; tumour; immune system; haematopoietic system;
KW reproductive system; liver; skeletal muscle; neurodegenerative disease;
KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;
KW blastocyst implantation; thrombosis; retinal disease;
KW retinal pigmentosis.
XX
OS Homo sapiens.
XX
PN WO200155172-A2.
XX
PD 02-AUG-2001.
XX
PF 26-JAN-2001; 2001WO-FR00253.
XX
PR 27-JAN-2000; 2000FR-0001035.
PR 12-OCT-2000; 2000FR-0013089.
XX
XX (FABR) FABRE MEDICAMENT SA PIERRE.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Elson G, Gauchat J, Plun-Favreau H, Chevallier S, Gascan H;
XX
XX WPI; 2001-488773/53.
DR N-PSDB; AAH74484.
XX
XX A complex comprising a NNT-1 protein and a CLF-1 and/or sCNTFRalpha
PT protein useful to treat neurodegenerative disease including Parkinson's
PT and Huntington's, obesity and cancer -
XX
XX Claim 2; Page 58; 67pp; French.
XX
XX The present sequence represents a human NNT-1 protein. The specification
CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or
CC sCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate
CC activity of the sCNTFRalpha/Gp130/LiFRbeta receptor complex, or to
CC induce phosphorylation of the tyrosine of Gp130 and LiFRbeta, in the
CC particularly where cells expressing the receptor complex are in the
CC central or peripheral nervous system, in neurons implicated in
CC neuro-muscular function or in skeletal muscle. The complex or
CC antibodies are also used to decrease the survival, growth or
CC proliferation of tumour cells or to facilitate the proliferation and/or
CC inhibit differentiation of cells stocks. The complex is also used to
CC modulate activity of the Gp130/LiFRbeta receptor or cells expressing
CC that receptor, particularly those cells implicated in the immune,
CC haematopoietic, nervous or reproductive system, the liver or skeletal
CC muscle. Molecules of the invention may be used to prevent or treat
CC neurodegenerative diseases including amyotrophic lateral sclerosis,
CC Parkinson's and Huntington's disease, to repair or regenerate nervous
CC or muscular tissue or to maintain muscular mass in paralysis patients.
CC They may also be used to treat cancer, obesity and associated diseases,
CC and to improve fertility, particularly to avoid endometriosis and/or

CC inhibitor to a patient. Also included are a method of diagnosing an
 CC IgE-related disease or susceptibility to an IgE-related disease, by
 CC determining the presence or amount of expression of an NNT1 polypeptide
 CC encoded by a NNT1 nucleotide sequence, its fragment or naturally
 CC occurring variant, and diagnosing an IgE-related disease or
 CC susceptibility of an IgE-related disease based on the presence or amount
 CC of expression of an NNT1 polypeptide and a pharmaceutical composition for use
 CC in treating IgE-related disease, comprising the NNT1 inhibitor.
 CC The NNT1 inhibitor is useful for preventing and treating IgE-related
 CC disease, modulating IgE levels, and treating allergic diseases e.g.
 CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,
 CC pollinosis, asthma, immune diseases and disorders, diseases involving
 CC abnormal cell proliferation including cancer, arteriosclerosis and
 CC immune system including rheumatoid arthritis, psoriatic arthritis,
 CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,
 CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
 CC bowel disease, transplant rejection, and graft versus host disease, and
 CC reproductive diseases and disorders including infertility, miscarriage,
 CC preterm labour and delivery, and endometriosis. The present sequence
 CC represents human NNT1.

XX Sequence 225 AA;

Query Match 100.0%; Score 1226; DB 23; Length 225;
 Best Local Similarity 100.0%; Pred. No. 1.4e-118;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGLACLTVMHLPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60
 DB 1 MDLRAGDSWGLACLTVMHLPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60

QY 61 YLYNLGPPNEPDPNPRIGAEPLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRL 120
 DB 61 YLYNLGPPNEPDPNPRIGAEPLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRL 120

QY 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPOLPGTPTWTGPAHSDFLQ 180
 DB 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPOLPGTPTWTGPAHSDFLQ 180

QY 181 KMDDFWLLKELQTLWRSKDFNRLKKQKQPPAAAVTLHGAHGF 225
 DB 181 KMDDFWLLKELQTLWRSKDFNRLKKQKQPPAAAVTLHGAHGF 225

RESULT 7

AAM25831
 ID AAM25831 standard; Protein; 253 AA.

XX AC AAM25831;

XX DT 16-OCT-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:1346.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnary; antidiabetic; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

OS Homo sapiens.

XX

PN WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-457603/49.

DR N-PSDB; AAM99772.

XX Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 20; Page 278; 1217pp; English.

XX AAM99166 to AAM99904 encode the human proteins given in AAM25225 to

CC AAM25963. The proteins can have activities based on the tissues and

CC cells they are expressed in, such as: antiinflammatory; antirheumatic;

CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;

CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;

CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;

CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;

CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;

CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

CC encoding them can be used in gene therapy, antitense therapy and vaccine

CC production. The proteins and polynucleotides are useful for screening for

CC agonists or antagonists of a protein and for the treatment and diagnosis

CC of disorders associated with the activity of a protein e.g. inflammation,

CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal

CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopenia, wounds, ulcers,

CC osteoporosis, severe combined immunodeficiency, eczema, allergic

CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

CC Alzheimer's disease, Parkinson's disease, neurodegenerative and

CC neurological disorders.

XX SQ Sequence 253 AA;

Query Match 100.0%; Score 1226; DB 22; Length 253;

Best Local Similarity 100.0%; Pred. No. 1.6e-118;

Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGLACLTVMHLPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 60

DB 29 MDLRAGDSWGLACLTVMHLPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGT 88

QY 61 YLYNLGPPNEPDPNPRIGAEPLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRL 120

DB 89 YLYNLGPPNEPDPNPRIGAEPLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRL 148

QY 121 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPOLPGTPTWTGPAHSDFLQ 180

DB 149 NRQATAELRRSLAHFCTSLQGLGSIAGVMAALGYPLPOLPGTPTWTGPAHSDFLQ 208

QY 181 KMDDFWLLKELQTLWRSKDFNRLKKQKQPPAAAVTLHGAHGF 225

DB 209 KMDDFWLLKELQTLWRSKDFNRLKKQKQPPAAAVTLHGAHGF 253

RESULT 8

AAE00828

ID AAE00828 standard; Protein; 223 AA.

XX AC AAE00828;

XX

immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity can be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 321 AA;

Query March 98.2%; Score 1204; DB 22; Length 321;
Best Local Similarity 99.1%; Pred. No. 4.2e-116;
Matches 221; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LRAGDSWGLACLTCLVWHLPAVPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGTYL 62
DB 99 LPTGDSWGLACLTCLVWHLPAVPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGTYL 158
QY 63 NYLGPFPNEPDPNPRLGAETLPRAATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 122
DB 159 NYLGPFPNEPDPNPRLGAETLPRAATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 218
QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQKM 182
DB 219 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQKM 278
QY 183 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225
DB 279 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 321

RESULT 10

AAM79399

ID AAM79399 standard; Protein; 321 AA.

XX AC AAM79399;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3045.

XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSB-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZN;
XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX DR WPI; 2001-476283/51.

DR N-PSDB; AAK52532.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX Claim 20; Page 237; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.

SQ Sequence 321 AA;

Query Match 98.2%; Score 1204; DB 22; Length 321;
Best Local Similarity 99.1%; Pred. No. 4.2e-116;
Matches 221; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LRAGDSWGLACLTCLVWHLPAVPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGTYL 62
DB 99 LPTGDSWGLACLTCLVWHLPAVPALNRTGDPGPGPSIQKTYDLYLHQLRSLAGTYL 158
QY 63 NYLGPFPNEPDPNPRLGAETLPRAATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 122
DB 159 NYLGPFPNEPDPNPRLGAETLPRAATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 218
QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQKM 182
DB 219 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQKM 278
QY 183 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225
DB 279 DDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 321

RESULT 11

AAM78415

ID AAM78415 standard; Protein; 260 AA.

XX AC AAM78415;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 1077.

XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.

Sun Feb 2 08:31:46 2003

PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR N-PSDB; AAK51548.
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 PT Claim 20; Page 3306; 6221pp; English.
 PS The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAW80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX Sequence 260 AA;
 PS Query Match 97.6%; Score 1197; DB 22; Length 260;
 CC Best Local Similarity 98.7%; Pred. No. 1.7e-115;
 CC Matches 220; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 LRAGDSWGLACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGTYL 62
 Db 38 LPTGDSWGLACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGTYL 97
 QY 63 NYLGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 122
 Db 98 NYLGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGLNR 157
 QY 123 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQRM 182
 Db 158 QAATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQRM 217
 QY 183 DDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225
 Db 218 DDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 260
 RESULT 12
 AAW29716 ID AAW29716 standard; Protein; 225 AA.
 XX AAW29716;
 XX 09-NOV-1998 (first entry)
 XX Mouse neurotrophic factor NNT-1.
 DE NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;
 XX Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
 KW peripheral neuropathy; dystrophy; neural retina degeneration;
 KW common variable immunodeficiency; CVID; selective IGA deficiency;
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
 KW therapy.

XX Mus sp.
 XX Key Location/Qualifiers
 FH 1..27 /label= Sig_peptide
 FT 28..225 /label= Mat_protein
 FT Protein
 XX WO9833922-A1.
 XX 06-AUG-1998.
 XX 02-FEB-1998; 98WO-US02363.
 XX 30-JAN-1998; 98US-0016534.
 PR 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 XX Chang M, Elliot GS, Sarmiento U, Senaldi G;
 WPI; 1998-437475/37.
 N-PSDB; AAW47512.
 XX Newly isolated nucleic acid encoding human or murine neurotrophic
 PT factor NNT-1 - useful for treatment of neurological and
 PT immunological diseases or inflammation, also as vaccine adjuvant
 XX Claim 13; Fig 5; 120pp; English.
 XX This is the amino acid sequence of a murine neurotrophic factor,
 CC designated NNT-1, that is a growth factor for neurons and for B or
 CC T cells. It was deduced from isolated NNT-1 cDNA (see AAW47512).
 CC Human NNT-1 (see AAW29715) is also provided. Vectors and host cells
 CC for use in the production of human murine recombinant NNT-1
 CC polypeptides. These are used to treat: (i) neurological or
 CC immunological diseases, specifically Alzheimer's, Parkinson's
 CC or Huntington's diseases, amyotrophic lateral sclerosis,
 CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
 CC degeneration of the neural retina, or conditions characterised by T
 CC or B cell defects, e.g. common variable immunodeficiency (CVID),
 CC selective IGA deficiency, hypogammaglobulinaemia and X-linked
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
 CC inflammation. NNT-1 is also able to boost immunoreactivity and
 CC antibody production following vaccination, and, since it inhibits
 CC tumour necrosis factor production, it may also be useful for
 CC treating sepsis. In addition, cells that have been engineered to
 CC express NNT-1 can be implanted, or nucleic acids are delivered in
 CC gene therapy vectors.
 XX Sequence 225 AA;
 PS Query Match 97.3%; Score 1193; DB 19; Length 225;
 CC Best Local Similarity 96.9%; Pred. No. 3.6e-115;
 CC Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGLACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGT 60
 Db 1 MDLRAGDSWGLACLTVMHLPAPVAPALNRTGDPGPGSIQKTYDLYLEHQLRSLAGT 60
 QY 61 YLYNLYGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGL 120
 Db 61 YLYNLYGPPNEPDPNPRGLAETLPATVDLEVWRSNDKRLTQNYEAYSHLLCYLRGL 120
 QY 121 NROQATAEELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
 Db 121 NROQATAEELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
 QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225
 Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKQPPAAAVTLHLGAHGF 225

```
RESULT 13
AAW56142
ID AAW56142 standard; Protein; 225 AA.
XX
AC AAW56142;
XX
DT 13-JUL-1998 (first entry)
XX
DE Amino acid sequence of murine neurotrophic factor NNT-1.
XX
KW Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /note= "signal peptide"
FT 28..225
FT /note= "mature peptide"
XX
PN US5741772-A.
XX
PD 21-APR-1998.
XX
PF 03-FEB-1997; 97US-0792019.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 1998-260526/23.
DR N-PSDB; AAW22654.
XX
PT Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
PT useful for stimulating growth of motor and sympathetic neurons
XX
PS Claim 2; Fig 5; 41pp; English.
XX
CC The present sequence represents a murine neurotrophic factor, designated
CC NNT-1, which is capable of stimulating growth of motor or sympathetic
CC neurons. The NNT-1 protein is useful in the treatment of neurological
CC diseases characterized by the degeneration and death of particular
CC classes of neurons. These diseases specifically include Parkinson's
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
CC stroke and various degenerative disorders affecting vision.
XX
SQ Sequence 225 AA;
Query Match 97.3%; Score 1193; DB 19; Length 225;
Best Local Similarity 96.9%; Pred. No. 3.6e-115;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MDLRAGDSWGLACLTIVLHLPVAPALNRTGDPGPGSIQKTYDLTRYLHQLRSLAGT 60
DB 1 MDLRAGDSWGLACLTIVLHLPVAPALNRTGDPGPGSIQKTYDLTRYLHQLRSLAGT 60
QY 61 YLNYLGPPFPNPPRLGATLPRATVDLEWVRSNDKLRLTQNYEAYSHLLCYLRGL 120
DB 61 YLNYLGPPFPNPPRLGATLPRATVNLVRSNDKLRLTQNYEAYSHLLCYLRGL 120
QY 121 NRQATLRLRSLAHFCTSLGGLGSLAGVMAALGYPLPQLPGTPTWTPGPAHSDFLQ 180
DB 121 NRQATLRLRSLAHFCTSLGGLGSLAGVMAALGYPLPQLPGTPTWTPGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWRSKADFNRLKKMQPPAAAVTLHLGAHGF 225
DB 181 KMDDFWLLKELQTLWRSKADFNRLKKMQPPAAAVTLHLGAHGF 225
```

```
RESULT 14
AAY87814
ID AAY87814 standard; Protein; 225 AA.
XX
AC AAY87814;
XX
DT 24-AUG-2000 (first entry)
XX
DE Murine NNT-1 protein.
XX
KW NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment;
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
KW retinopathy; immune disorder; hematopoietic disorder.
XX
OS Mus sp.
XX
PN US6054294-A.
XX
PD 25-APR-2000.
XX
PF 12-DEC-1997; 97US-0988819.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 2000-338492/29.
DR N-PSDB; AAA39483.
XX
PT New nucleic acids encoding neurotrophic factors useful for stimulating
PT growth of motor or sympathetic neurons for treating neuron cell damage
XX
PS Claim 2b; Fig 5; 42pp; English.
XX
CC This invention describes a novel nucleic acid molecule (I) encoding a
CC novel neurotrophic factor (NNT-1) (II) which has neurotropic,
CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
CC ophthalmological activity. (I) is useful for producing NNT-1
CC polypeptides which are useful for treating patients in whom various
CC cells of the central, autonomic, or peripheral nervous system have
CC degenerated and/or have been damaged by congenital disease, trauma,
CC mechanical damage, surgery, stroke, ischemia, infection, metabolic
CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
CC proteins are used to treat diseases like Alzheimer's, Parkinson's,
CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
CC disease, peripheral neuropathy induced by diabetes or other metabolic
CC disorders, and/or dystrophies or degeneration of the neural retina such
CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
CC night blindness, progressive cone-rod degeneration, immune disorders and
CC hematopoietic disorders. (I) is effective in treating neurological
CC conditions and promotes neuron regeneration. Neural functions are
CC effectively restored in patients suffering from various neurological
CC disorders. This sequence represents the murine NNT-1 protein described in
CC the method of the invention.
XX
SQ Sequence 225 AA;
Query Match 97.3%; Score 1193; DB 21; Length 225;
Best Local Similarity 96.9%; Pred. No. 3.6e-115;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MDLRAGDSWGLACLTIVLHLPVAPALNRTGDPGPGSIQKTYDLTRYLHQLRSLAGT 60
DB 1 MDLRAGDSWGLACLTIVLHLPVAPALNRTGDPGPGSIQKTYDLTRYLHQLRSLAGT 60
QY 61 YLNYLGPPFPNPPRLGATLPRATVDLEWVRSNDKLRLTQNYEAYSHLLCYLRGL 120
```


CC immune system including rheumatoid arthritis, psoriatic arthritis,
CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,
CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
CC bowel disease, transplant rejection, and graft versus host disease, and
CC reproductive diseases and disorders including infertility, miscarriage,
CC preterm labour and delivery, and endometriosis. The present sequence
CC represents Mouse NNT1.
XX
SQ Sequence 225 AA;
Query Match 97.3%; Score 1193; DB 23; Length 225;
Best Local Similarity 96.9%; Pred. No. 3.6e-115;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLTRYLHQLSLAGT 60
DB 1 MDLRAGDSWGMGLACTVLMHLPVAPALNRTGDPGPGPSIQKTYDLYLTRYLHQLSLAGT 60
QY 61 YLNYLGPPFNEPDPNPPRLGAETLPRATVLDLVWESLNDKRLTQNYEAYSHLLCYLRGL 120
DB 61 YLNYLGPPFNEPDPNPPRLGAETLPRATVLDLVWESLNDKRLTQNYEAYSHLLCYLRGL 120
QY 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTGPAHSDFLQ 180
DB 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225
DB 181 KMDDFWLLKELQTLWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225

Search completed: January 27, 2003, 15:41:34
Job time : 35 secs

DB 61 YLNYLGPPFNEPDPNPPRLGAETLPRATVLDLVWESLNDKRLTQNYEAYSHLLCYLRGL 120
QY 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTGPAHSDFLQ 180
DB 121 NRQATAELRESLAHFCTSLQGLLSIAGVMAALGYPLPQPLGTEPTWTGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225
DB 181 KMDDFWLLKELQTLWLRSAKDFNRLKKKQPPAAAVTLHLGAHGF 225
RESULT 15
ID AAU78177 standard; Protein; 225 AA.
XX
AC AAU78177;
XX
DT 05-JUN-2002 (first entry)
XX
DE Mouse novel neurotrophic factor NNT1.
XX
KW Mouse; NNT1; neurotrophic factor; Ige-related disease;
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
KW inflammatory bowel disease; transplant rejection; reproductive disorder;
KW graft versus host disease; infertility; miscarriage; preterm labour.
XX
OS Mus sp.
XX
PN WO200215977-A2.
XX
PD 28-FEB-2002.
XX
PF 17-AUG-2001; 2001WO-US25906.
XX
PR 19-AUG-2000; 2000US-226436P.
PR 16-AUG-2001; 2001US-0931704.
XX
PA (AMGE-) AMGEN INC.
XX
PI Senaldi G;
XX
DR WPI; 2002-280867/32.
DR N-PSDB; ABK11649.
XX
PT Treating Immunoglobulin E-related disease, modulating Ige levels in a
PT patient, preventing Ige-related disease and treating allergic diseases,
PT involves administering NNT-1 inhibitor to a patient -
XX
XX Claim 2; Fig 5; 63pp; English.

CC The invention relates to treating Immunoglobulin E (Ige)-related disease,
CC modulating Ige levels in a patient, preventing an Ige-related disease,
CC and treating allergic diseases, comprising administering a
CC therapeutically effective amount of novel neurotrophic factor (NNT)-1
CC inhibitor to a patient. Also included are a method of diagnosing an
CC Ige-related disease or susceptibility to an Ige-related disease, by
CC determining the presence or amount of expression of an NNT1 polypeptide
CC encoded by a NNT1 nucleotide sequence, its fragment or naturally
CC occurring variant, and diagnosing an Ige-related disease or
CC susceptibility of an Ige-related disease based on the presence or amount
CC of expression of the polypeptide and a pharmaceutical composition for use
CC in treating Ige-related disease, comprising the NNT1 inhibitor.
CC The NNT1 inhibitor is useful for preventing and treating Ige-related
CC disease, modulating Ige levels, and treating allergic diseases e.g.
CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,
CC pollinosis, asthma, immune diseases and disorders, diseases involving
CC abnormal cell proliferation including cancer, arteriosclerosis and
CC vascular restenosis, diseases and conditions relating to dysfunction of

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:42:42 ; Search time 9 Seconds
(without alignments)
504.464 Million cell updates/sec

Title: US-09-931-704-2
Perfect score: 1226
Sequence: 1 MDLRAGDSWGLACLCTVLW.....KKKQPPAAAVTLHLGAHGF 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues
Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1226	100.0	225	10	US-09-931-704-2 Sequence 2, Appli
2	1193	97.3	225	10	US-09-931-704-5 Sequence 5, Appli
3	885	72.2	163	10	US-09-864-761-40014 Sequence 40014, A
4	118.5	9.7	203	10	US-09-896-856-3 Sequence 3, Appli
5	96.5	7.9	243	9	US-10-000-776-6 Sequence 6, Appli
6	96.5	7.9	243	9	US-09-791-497-8 Sequence 8, Appli
7	96.5	7.9	243	10	US-09-810-052-5 Sequence 5, Appli
8	91.5	7.5	201	10	US-09-901-540-3 Sequence 3, Appli
9	91.5	7.5	201	10	US-09-896-856-8 Sequence 8, Appli
10	91.5	7.5	201	10	US-09-901-257-3 Sequence 3, Appli
11	91	7.4	195	10	US-09-770-361-5 Sequence 5, Appli
12	91	7.4	242	9	US-10-000-776-2 Sequence 2, Appli
13	91	7.4	242	9	US-09-791-497-2 Sequence 2, Appli
14	90	7.3	232	10	US-09-810-052-2 Sequence 2, Appli
15	83	6.8	625	10	US-09-771-161A-242 Sequence 242, App
16	83	6.8	625	10	US-09-771-161A-243 Sequence 243, App
17	82	6.7	200	10	US-09-770-361-8 Sequence 8, Appli
18	81.5	6.6	218	10	US-09-893-737-28 Sequence 28, Appli
19	81	6.6	200	10	US-09-770-361-6 Sequence 6, Appli

Sequence 4, Appli
Sequence 12, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 10, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 10, Appli
Sequence 16, Appli
Sequence 16, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 8, Appli
Sequence 10, Appli
Sequence 60, Appli
Sequence 2, Appli
Sequence 1344, Ap
Sequence 3, Appli
Sequence 5603, Ap
Sequence 12181, A
Sequence 12995, A
Sequence 13148, A

20 80.5 6.6 1399 9 US-09-388-221-4
21 80.5 6.6 1424 9 US-09-388-221-12
22 80.5 6.6 1429 10 US-09-996-617-2
23 80.5 6.6 1429 10 US-09-931-071-2
24 80.5 6.6 1443 9 US-09-388-221-6
25 80.5 6.6 1454 9 US-09-388-221-10
26 80.5 6.6 1473 9 US-09-388-221-2
27 79.5 6.5 348 10 US-09-730-617-2
28 79 6.4 418 10 US-09-946-807-3
29 79 6.4 418 10 US-09-795-668-3
30 79 6.4 418 10 US-09-795-686-3
31 77.5 6.3 200 10 US-09-770-361-10
32 77.5 6.3 379 9 US-09-860-846-16
33 77.5 6.3 379 10 US-09-861-289-16
34 77.5 6.3 3782 9 US-09-860-846-4
35 77.5 6.3 3782 10 US-09-861-289-4
36 77 6.3 234 9 US-10-000-776-8
37 77 6.3 234 9 US-09-791-497-10
38 77 6.3 439 12 US-10-078-929-60
39 76 6.2 416 9 US-10-124-429-2
40 76 6.2 451 10 US-09-764-864-1344
41 75 6.1 199 10 US-09-770-361-3
42 75 6.1 917 10 US-09-815-242-5603
43 75 6.1 920 10 US-09-815-242-12181
44 75 6.1 920 10 US-09-815-242-12995
45 75 6.1 920 10 US-09-815-242-13148

ALIGNMENTS

RESULT 1

US-09-931-704-2
; Sequence 2, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating IGE-Related Disease Using
; TITLE OF INVENTION: Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-931-704-2

Query Match 100.0%; Score 1226; DB 10; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.5e-113;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGLACLCTVLWHLPAVPALNRTGDPGPGSIQKTYDLTRYLHQLRSLAGT 60
Db 1 MDLRAGDSWGLACLCTVLWHLPAVPALNRTGDPGPGSIQKTYDLTRYLHQLRSLAGT 60

QY 61 YLNYLGPPFPNPPRLGAETLPRATVDLVWRSNDKLRLLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFPNPPRLGAETLPRATVDLVWRSNDKLRLLTQNYEAYSHLLCYLRGL 120

QY 121 NRQATAELRSLAHFTCTSLQGLGSTAGVMAALGYPLPQPLPCTETWTGPAHSDFLQ 180
Db 121 NRQATAELRSLAHFTCTSLQGLGSTAGVMAALGYPLPQPLPCTETWTGPAHSDFLQ 180

QY 181 KMDDFWLLKELQTLWRSKDFNLRKKKQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWRSKDFNLRKKKQPPAAAVTLHLGAHGF 225

RESULT 2

US-09-931-704-5

Sequence 5, Application US/09931704

Patent No. US20020041873A1

GENERAL INFORMATION:

APPLICANT: Senaldi, Giorgio

TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using N

FILE REFERENCE: A-695

CURRENT APPLICATION NUMBER: US/09/931,704

CURRENT FILING DATE: 2001-08-16

PRIOR APPLICATION NUMBER: US 60/226,436

PRIOR FILING DATE: 2000-08-18

NUMBER OF SEQ ID NOS: 5

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 225

TYPE: PRT

ORGANISM: Murine

US-09-931-704-5

Query Match 97.3%; Score 1193; DB 10; Length 225;

Best Local Similarity 96.9%; Pred. No. 1.1e-109;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMLACCTVLVHLPAVPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

Db 1 MDLRAGDSWGMLACCTVLVHLPAVPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPNEPDPNPRLAGETLPRATVDLEVRSLNDKURLTQNYEAYSHLLCYLRGL 120

Db 61 YLNYLGPPNEPDPNPRLAGETLPRATVNLVWRSLNDRLRLTQNYEAYSHLLCYLRGL 120

Qy 121 NQQAATAEURLRSIAHFTCSLQGLGSGTAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180

Db 121 NQQAATAEURLRSIAHFTCSLQGLGSGTAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSASDKDFNRLKKKMQPPAAAATLHLGAHGF 225

Db 181 KMDDFWLLKELQTLWRSASDKDFNRLKKKMQPPAAAATLHLGAHGF 225

RESULT 3

US-09-864-761-40014

Sequence 40014, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Acomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669


```
QY 47 TRYLEHQLRSLAGTYLN-----YLGPPFNEPDNPPRLGAETLPRAVDLVWVRSLN 98
Db 54 ARKLSEVRGQHRFAESHLPGVNLVLP-----LG-EQLPDVSLTFQAWRRLS 101
QY 99 DKRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLLSIAGVMAALGYP 157
Db 102 DPERLCFISTTLQPPHAPLGGLTQGRWTNMMERQMLWAMRLDLRDLQRLHRLFQVLAAGEN 161
QY 158 LPQP-----LPQTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDF 202
Db 162 LPPEEEEEEEEEERKGLLPGALGALQGPQVSWPQLLSTYRLLSLHSLVLSRAVREL 221
QY 203 NRLKK 207
Db 222 LLLSK 226

RESULT 7
US-09-810-052-5
; Sequence 5, Application US/09810052
; Patent No. US20020009775A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Preenell, Scott R.
; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51
; FILE REFERENCE: 00-24
; CURRENT APPLICATION NUMBER: US/09/810,052
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/190,410
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/199,443
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-810-052-5

Query Match 7.9%; Score 96.5; DB 10; Length 243;
Best Local Similarity 23.7%; Pred. No. 0.04;
Matches 58; Conservative 26; Mismatches 104; Indels 57; Gaps 9;

QY 1 MDLRAGD-SWGMACLCITVL-----WHLPVAPALNRTGDPGPGPSIQK-----TYDL 46
Db 1 MGQTAGDLGWRISLALLLLVQAGVWGFRPPG-----RPQLSLQELRREFTVLSHL 53
QY 47 TRYLEHQLRSLAGTYLN-----YLGPPFNEPDNPPRLGAETLPRAVDLVWVRSLN 98
Db 54 ARKLSEVRGQHRFAESHLPGVNLVLP-----LG-EQLPDVSLTFQAWRRLS 101
QY 99 DKRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLLSIAGVMAALGYP 157
Db 102 DPERLCFISTTLQPPHAPLGGLTQGRWTNMMERQMLWAMRLDLRDLQRLHRLFQVLAAGEN 161
QY 158 LPQP-----LPQTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDF 202
Db 162 LPPEEEEEEEEEERKGLLPGALGALQGPQVSWPQLLSTYRLLSLHSLVLSRAVREL 221
QY 203 NRLKK 207
Db 222 LLLSK 226

RESULT 8
US-09-901-540-3
; Sequence 3, Application US/09901540
; Patent No. US20020102622A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
```

```
; APPLICANT: Lawrence, David A.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret A.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Cardiostrophin-1 Compositions and Methods for the
; TITLE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P253D2
; CURRENT APPLICATION NUMBER: US/09/901,540
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 09/648,258
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 09/033,114
; PRIOR FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: US 08/733,850
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: US 08/443,129
; PRIOR FILING DATE: 1995-05-17
; PRIOR APPLICATION NUMBER: US 08/286,304
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: US 08/233,609
; PRIOR FILING DATE: 1994-04-25
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Human
US-09-901-540-3

Query Match 7.5%; Score 91.5; DB 10; Length 201;
Best Local Similarity 25.9%; Pred. No. 0.096;
Matches 45; Conservative 24; Mismatches 92; Indels 13; Gaps 5;

QY 40 IQKTYDITRYLEHQLRSLAGTYLNLYLGPPFNEPDNPPRLGAETLPRAVDLVWVRSLN 99
Db 27 IROTHSLAHLTKYAEQLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 55
QY 100 KLRLTONVEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLLSIAGVMAAL 154
Db 86 RLRL--DAALAAALPPLDVAQCRAELNPRAPRLRLLEDAARQARALGAVALAAL 143
QY 155 GYPLPQLPGTPE--TWTPGAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRL 205
Db 144 G--AANRGPRAEPPAATAASAATGVPFAKVLGLRVCGLYREWLSRTGDLQQL 195

RESULT 9
US-09-896-856-8
; Sequence 8, Application US/09896856
; Patent No. US20020137189A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; Chien, Kenneth
; King, Kathleen
; Pennica, Diane
; Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/896,856
```

```
; FILING DATE: 29-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER: US 08/733,850
; FILING DATE: 18-OCT-1996
; APPLICATION NUMBER: US 08/471,112
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: 08/233,609
; FILING DATE: 25-APR-1994
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: 08/443129
; FILING DATE: 17-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P0894PID2C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2066
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-896-856-8

Query Match          7.5%; Score 91.5; DB 10; Length 201;
Best Local Similarity 25.9%; Pred. No. 0.096;
Matches 45; Conservative 24; Mismatches 92; Indels 13; Gaps 5;

QY 40 IQKTYDLTRYLEHQRLSLAGTYLNYLGGPPFNEPDPNPPRLGAETLPRAIVDLEWVRSND 99
DB 27 IRQTHSLAHLTKYAEQLQEVVQLQGGDFGLPSFSPRLPVAGL-SAPAPSHAGLPVHE 85
QY 100 KLRLTONYEAYSHLLCYLRGLNRQAA-----TAELRRLSLAHFCTSLQGLLSIAGVMAAL 154
DB 86 RLRL--DAAALAAPLLDVCRCRQAEINPRAPRLRLRLLEDAARQARALGAAVEALLAAL 143
QY 155 GYPLPQPLPGTEP---TWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRL 205
DB 144 G--AANRGFRAEPPTAASATGVFFPAKVLGLRVCGLYREWLSTRTEGDLGQL 195

RESULT 10
US-09-901-257-3
; Sequence 3, Application US/09901257
; Patent No. US20020146707A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
; APPLICANT: Lawrence, David A.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret A.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the
; FILE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P2533D3
; CURRENT APPLICATION NUMBER: US/09/901,257
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 09/648,252
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 09/033,114
; PRIOR FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: US 08/733,850
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: US 08/443,129
; PRIOR FILING DATE: 1995-05-17
; PRIOR APPLICATION NUMBER: US 08/286,304
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: US 08/233,609
```

```
; PRIOR FILING DATE: 1994-04-25
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Human
; US-09-901-257-3

Query Match          7.5%; Score 91.5; DB 10; Length 201;
Best Local Similarity 25.9%; Pred. No. 0.096;
Matches 45; Conservative 24; Mismatches 92; Indels 13; Gaps 5;

QY 40 IQKTYDLTRYLEHQRLSLAGTYLNYLGGPPFNEPDPNPPRLGAETLPRAIVDLEWVRSND 99
DB 27 IRQTHSLAHLTKYAEQLQEVVQLQGGDFGLPSFSPRLPVAGL-SAPAPSHAGLPVHE 85
QY 100 KLRLTONYEAYSHLLCYLRGLNRQAA-----TAELRRLSLAHFCTSLQGLLSIAGVMAAL 154
DB 86 RLRL--DAAALAAPLLDVCRCRQAEINPRAPRLRLRLLEDAARQARALGAAVEALLAAL 143
QY 155 GYPLPQPLPGTEP---TWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRL 205
DB 144 G--AANRGFRAEPPTAASATGVFFPAKVLGLRVCGLYREWLSTRTEGDLGQL 195

RESULT 11
US-09-770-361-5
; Sequence 5, Application US/09770361
; Patent No. US20020123462A1
; GENERAL INFORMATION:
; APPLICANT: Fandl, James
; APPLICANT: Stahl, Neil
; TITLE OF INVENTION: Modified Ciliary Neurotrophic Factor, Method Of Making
; FILE OF INVENTION: And Methods Of Use Thereof
; FILE REFERENCE: REG142
; CURRENT APPLICATION NUMBER: US/09/770,361
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 09/031,693
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 08/308,736
; PRIOR FILING DATE: 1994-09-19
; PRIOR APPLICATION NUMBER: 07/959,284
; PRIOR FILING DATE: 1992-10-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 195
; TYPE: PRT
; ORGANISM: CHICKEN
; US-09-770-361-5
```

```
Query Match          7.4%; Score 91; DB 10; Length 195;
Best Local Similarity 27.3%; Pred. No. 0.1;
Matches 51; Conservative 21; Mismatches 83; Indels 32; Gaps 9;

QY 46 LTRYLEHQRLSLAGTYLNYLGGPPFNEPDPNPPRLGAETLPRAIVDLEWVRSNDKLRUTQ 105
DB 23 LARKMRSDVTDLLDIYVERQG-----LDASISVAADVGPPTAAV--ERWAEQTGTQRLLD 75
QY 106 N---YEAYSHLLCYLRGLNRQAA---ATAELRRSLA-----HFCTSIQGLLSIAGVMA 152
DB 76 NLAAAYRAFRITLLAQMLEEQEELLGDTDAELGPAALAMLLQVSFAFYHLEELL-----ELE 130
QY 153 ALGYPLPQPLPGTEPTWTPGAH--SDFLQKMDDFWLLKELQTLWRSKDFNRLKKMQP 211
DB 131 SRGAPAE---GSEP---PAPPRLSLFEQKLRLGLRVRLRELAQWAVRSVRDLRLQLSKHGP 184
QY 212 PAAAVTL 218
DB 185 SGAALGL 191
```

RESULT 12

US-10-000-776-2

; Sequence 2, Application US/10000776

; Patent No. US20020164609A1

; GENERAL INFORMATION:

; APPLICANT: Timans, Jacqueline C.

; APPLICANT: Pflanz, Stefan K.-H.

; APPLICANT: Kastelein, Robert A.

; APPLICANT: Bazan, Jose F.

; APPLICANT: Rennick, Donna

; APPLICANT: de Waal Malefyt, Rene

; APPLICANT: Cheung, Jeanne

; TITLE OF INVENTION: MAMMALIAN CYTOKINES; RELATED REAGENTS

; FILE REFERENCE: DX01040K3

; CURRENT APPLICATION NUMBER: US/10/000,776

; CURRENT FILING DATE: 2001-11-30

; PRIOR APPLICATION NUMBER: 09/791,497

; PRIOR FILING DATE: 2001-02-22

; PRIOR APPLICATION NUMBER: 09/627,897

; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/146,581

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/147,763

; PRIOR FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 2

; LENGTH: 242

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-000-776-2

Query Match 7.4%; Score 91; DB 9; Length 242;

Best Local Similarity 23.0%; Pred. No. 0.14;

Matches 56; Conservative 26; Mismatches 105; Indels 56; Gaps 8;

QY 2 DLKAGDSWGLACLCCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDLTR 48
DB 2 DLENNPKIGLSLLPLLLVQAGVWGFPFPG-----RQLSLQELRRFTVSLHLAR 54
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRPGAETLPRATVDLEWVRSNDK 100
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRRLSDP 102
QY 101 LRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLGSIAGVMAALGYPLP 159
DB 103 ERLCFISTTLQPFHAPLGGLGTQGRWTNMERMQLWAMELDLRLDLRHLRFQVLAAGFNLP 162
QY 160 QP-----LPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204
DB 163 EEEEEEEEEERKGLLPALGALQPAQVSWPQLLSTYRLHLSLELVLSRAVRELLL 222
QY 205 LKK 207
DB 223 LSK 225

RESULT 13

US-09-791-497-2

; Sequence 2, Application US/09791497

; Publication No. US20030008343A1

; GENERAL INFORMATION:

; APPLICANT: Timans, Jacqueline C.

; APPLICANT: Kastelein, Robert A.

; APPLICANT: Bazan, J. Fernando

; APPLICANT: Pflanz, Stefan

; TITLE OF INVENTION: Mammalian Cytokines; Related Reagents

; FILE REFERENCE: DX01040K2

; CURRENT APPLICATION NUMBER: US/09/791,497

; CURRENT FILING DATE: 2001-02-22

; PRIOR APPLICATION NUMBER: 09/627,897

; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/146,581

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/147,763

; PRIOR FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 2

; LENGTH: 242

; TYPE: PRT

; ORGANISM: primate; surmised Homo sapiens

; US-09-791-497-2

Query Match

7.4%; Score 91; DB 9; Length 242;

Best Local Similarity 23.0%; Pred. No. 0.14;

Matches 56; Conservative 26; Mismatches 105; Indels 56; Gaps 8;

QY 2 DLKAGDSWGLACLCCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDLTR 48
DB 2 DLENNPKIGLSLLPLLLVQAGVWGFPFPG-----RQLSLQELRRFTVSLHLAR 54
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRPGAETLPRATVDLEWVRSNDK 100
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRRLSDP 102
QY 101 LRLTQNYEAYSHLLCYLRGLNRQAATAELRR-SLAHFCTSLQGLGSIAGVMAALGYPLP 159
DB 103 ERLCFISTTLQPFHAPLGGLGTQGRWTNMERMQLWAMELDLRLDLRHLRFQVLAAGFNLP 162
QY 160 QP-----LPGTEPTWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204
DB 163 EEEEEEEEEERKGLLPALGALQPAQVSWPQLLSTYRLHLSLELVLSRAVRELLL 222
QY 205 LKK 207
DB 223 LSK 225

RESULT 14

US-09-810-052-2

; Sequence 2, Application US/09810052

; Patent No. US20020009775A1

; GENERAL INFORMATION:

; APPLICANT: Conklin, Darrell C.

; APPLICANT: Presnell, Scott R.

; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51

; FILE REFERENCE: 00-24

; CURRENT APPLICATION NUMBER: US/09/810,052

; CURRENT FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: 60/190,410

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/199,443

; PRIOR FILING DATE: 2000-04-25

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 232

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-09-810-052-2

Query Match

7.3%; Score 90; DB 10; Length 232;

Best Local Similarity 23.2%; Pred. No. 0.16;

Matches 51; Conservative 25; Mismatches 94; Indels 50; Gaps 7;

QY 19 LWHLPVAPALNRTGDPGPGPSIQK-----TYDLTRYLEHQLRSLAGTYLN----- 63
DB 15 VNGFPRPG-----RQLSLQELRRFTVSLHLARLKLSEVRGQAHRAESHLPGVNL 67
QY 64 YLGPPFNEPDPNPRPGAETLPRATVDLEWVRSNDKRLTQNYEAYSHLLCYLRGLNRQ 123
DB 68 YLLP-----LG-EQLPDVSLTFQAWRRLSDPERLCFTSTTLQPFHAPLGGLGTQ 115
QY 124 AATAELRR-SLAHFCTSLQGLGSIAGVMAALGYPLPQP-----LPQTEP 167

Db 116 GRWTMMERQMLWAMRLDRLDQLRHFRFOVLAAGFNLPPEEEEEEEERKGLLPALG 175
Qy 168 TWTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLKK 207
Db 176 SALQSPAQVSWPQLLSTYRLLSLELVLSRAVRELLLSK 215

RESULT 15

US-09-771-161A-242
; Sequence 242, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771.161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 242
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-242

Query Match 6.8%; Score 83; DB 10; Length 625;
Best Local Similarity 25.2%; Pred. No. 2.8;
Matches 60; Conservative 26; Mismatches 78; Indels 74; Gaps 15;
Qy 2 DLKAGDSGMLACLCTVLWHLPAV-----PALNRTGDPGPGPSIQKTYDLYLEHQLR 55
Db 382 DIREGAPLGV-----HLFFVGYSYSCMALRDSVPGPTP-----MELEAE 421
Qy 56 SLAGTYLNYLGLPPFNEPDNPPRLGAE-----TLPRATVDLEV-WRSINDKLR---LTQN 106
Db 422 QLLEPHVQ---APSLPSVSPQDETAEVAVPAAPAAEAEEVTLRELQEPLEEEVLTR- 477
Qy 107 YEAYSHLLCYLRGNLQRAAT----AEUR-RSLAHFCTSLQGLLGI--AGVMAALGYPLP 159
Db 478 -QSLREMEAIRTDNQNFASQLREAEARNRDLAEHVRQLQERMELLQAEAGATAVIGVPS 536
Qy 160 QPLFGTEPTWTPGAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAAVT 217
Db 537 R---ATDP-----PSH-----MAPRPLWASARWVG-----QAPCTAAT 567

Search completed: January 27, 2003, 15:49:54
Job time : 10 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:22 ; Search time 13 Seconds
(without alignments)
509.243 Million cell updates/sec

Title: US-09-931-704-2

Perfect score: 1226

Sequence: 1 MDLRAGDSWGMCLCTVLM.....KKKMQPPAAAVTLHGAGHF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1226	100.0	225	1 US-08-792-019B-2	Sequence 2, Appli
2	1226	100.0	225	3 US-08-106-182-2	Sequence 2, Appli
3	1226	100.0	225	3 US-08-988-819-2	Sequence 2, Appli
4	1226	100.0	225	4 US-08-016-534-2	Sequence 2, Appli
5	1193	97.3	225	1 US-08-792-019B-5	Sequence 5, Appli
6	1193	97.3	225	3 US-08-988-819-5	Sequence 5, Appli
7	1193	97.3	225	4 US-08-016-534-5	Sequence 5, Appli
8	124.5	10.2	203	3 US-08-106-182-3	Sequence 3, Appli
9	118.5	9.7	203	1 US-08-233-609-3	Sequence 3, Appli
10	118.5	9.7	203	1 US-08-444-083-3	Sequence 3, Appli
11	118.5	9.7	203	1 US-08-286-304-3	Sequence 3, Appli
12	118.5	9.7	203	1 US-08-442-745-3	Sequence 3, Appli
13	118.5	9.7	203	1 US-08-443-129-3	Sequence 3, Appli
14	118.5	9.7	203	1 US-08-443-952-3	Sequence 3, Appli
15	118.5	9.7	203	1 US-08-443-130-3	Sequence 3, Appli
16	118.5	9.7	203	3 PCT-US95-04467-3	Sequence 3, Appli
17	118.5	9.7	203	1 US-08-898-911-3	Sequence 3, Appli
18	91.5	7.5	201	1 US-08-444-083-8	Sequence 8, Appli
19	91.5	7.5	201	1 US-08-286-304-8	Sequence 8, Appli
20	91.5	7.5	201	1 US-08-442-745-8	Sequence 8, Appli
21	91.5	7.5	201	1 US-08-443-129-8	Sequence 8, Appli
22	91.5	7.5	201	1 US-08-443-952-8	Sequence 8, Appli
23	91.5	7.5	201	1 US-08-443-130-8	Sequence 8, Appli
24	91.5	7.5	201	1 US-08-792-019B-11	Sequence 11, Appli
25	91.5	7.5	201	3 US-08-106-182-4	Sequence 4, Appli
26	91.5	7.5	201	3 US-08-988-819-11	Sequence 11, Appli
27	91.5	7.5	201	3 US-08-898-911-8	Sequence 8, Appli

28 91.5 7.5 201 4 US-09-016-534-11 Sequence 11, Appli

29 91.5 7.5 201 5 PCT-US95-04467-8 Sequence 8, Appli

30 91.5 7.5 1182 4 US-09-287-354-6 Sequence 6, Appli

31 91 7.4 195 1 US-07-959-284-5 Sequence 5, Appli

32 91 7.4 195 2 US-08-308-736A-5 Sequence 5, Appli

33 91 7.4 195 4 US-08-645-107A-5 Sequence 5, Appli

34 91 7.4 195 4 US-09-197-349-5 Sequence 2, Appli

35 91 7.4 195 5 PCT-US92-08258-2 Sequence 2, Appli

36 91 7.4 195 5 PCT-US93-09649A-5 Sequence 5, Appli

37 91 7.4 195 5 PCT-US93-09649-5 Sequence 5, Appli

38 87.5 7.1 1207 4 US-09-287-354-5 Sequence 5, Appli

39 85.5 7.0 560 2 US-08-756-317-11 Sequence 11, Appli

40 84 6.9 200 4 US-08-949-155-4 Sequence 4, Appli

41 84 6.9 200 4 US-09-819-964-4 Sequence 4, Appli

42 82.5 6.7 582 2 US-08-422-699A-9 Sequence 9, Appli

43 82.5 6.7 582 2 US-08-422-706B-9 Sequence 9, Appli

44 82.5 6.7 984 4 US-09-287-354-2 Sequence 2, Appli

45 82.5 6.7 1189 4 US-09-287-354-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-792-019B-2
; Sequence 2, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-792-019B-2

Query Match 100.0%; Score 1226; DB 1; Length 225;

Best Local Similarity 100.0%; Pred. No. 3.7e-129;

Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMCLCTVLMHLPVAPALNRTGDPGPGSIQKTYDLYLRYLHQLRSLAGT 60

Db 1 MDLRAGDSWGMCLCTVLMHLPVAPALNRTGDPGPGSIQKTYDLYLRYLHQLRSLAGT 60

Qy 61 YLYNLGPPFPNPPRLGAETLPVATVLEWRSNDKRLRLTONYEAYSHLLCYLRGL 120

Db 61 YLYNLGPPFPNPPRLGAETLPVATVLEWRSNDKRLRLTONYEAYSHLLCYLRGL 120

Qy 121 NRQATAEARRSLAHFCTSLQGLGSIAGVMAALGYPLQPLGTEPTTTPGPAHSDFLQ 180

Db 121 NRQATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180
Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225

RESULT 2
US-09-106-182-2
; Sequence 2, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-106-182-2

Query Match 100.0%; Score 1226; DB 3; Length 225;
Best Local Similarity 100.0%; Pred. No. 3.7e-129;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMLACLTCLVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMLACLTCLVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPFNEPDPNPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFNEPDPNPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180
Db 121 NRQATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225

RESULT 3
US-08-988-819-2

; Sequence 2, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-988-819-2

Query Match 100.0%; Score 1226; DB 3; Length 225;
Best Local Similarity 100.0%; Pred. No. 3.7e-129;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMLACLTCLVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGMLACLTCLVWLHPVAPALNRTGDPGPGSIQKTYDLTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPFNEPDPNPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFNEPDPNPRLGAETLPRATVDLEWRSNDKRLTQNYEAYSHLLCYLRGL 120

Qy 121 NRQATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180
Db 121 NRQATAELRRSLAHFCTSLQGLLSIAGVMAALGYPLPQPLPGTPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAATLHLGAHGF 225

RESULT 4
US-09-016-534-2
; Sequence 2, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS

STATE: CA
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,534
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/792,019
FILING DATE: 03-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: COOK, ROBERT R.
REGISTRATION NUMBER: 31,602
REFERENCE/DOCKET NUMBER: A-442B
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-016-534-2

Query Match 100.0%; Score 1226; DB 4; Length 225;
Best Local Similarity 100.0%; Pred. No. 3.7e-129;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACICTVLVHLPVAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGLACICTVLVHLPVAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGT 60
Qy 61 YNVLGPPFNEPDPFPRLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Db 61 YNVLGPPFNEPDPFPRLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Qy 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
Db 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKXQPPAAAVTILHGAHF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKXQPPAAAVTILHGAHF 225

RESULT 5
US-08-792-019B-5
Sequence 5, Application US/08792019B
Patent No. 5741772
GENERAL INFORMATION:
APPLICANT: CHANG, MING-SHI
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC.
STREET: 1840 DEHAVILLAND DRIVE
CITY: THOUSAND OAKS
STATE: CA
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/792,019B
FILING DATE: 03-FEB-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:

NAME: COOK, ROBERT R.
REGISTRATION NUMBER: 31,602
REFERENCE/DOCKET NUMBER: A-442
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-792-019B-5

Query Match 97.3%; Score 1193; DB 1; Length 225;
Best Local Similarity 96.9%; Pred. No. 1.8e-125;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACICTVLVHLPVAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGLACICTVLVHLPVAVPALNRTGDPGPGPSIQKTYDLTRYLEHQLRSLAGT 60
Qy 61 YNVLGPPFNEPDPFPRLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Db 61 YNVLGPPFNEPDPFPRLGAETLPRAVDLEWVRSNDKRLTQNYEAYSHLLCYLRGL 120
Qy 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
Db 121 NRQAATAEIRSLAHFCTSLQGLGSIAGVMAALGYPLPQPLPGTEPTWTPGPAHSDFLQ 180
Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKXQPPAAAVTILHGAHF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKXQPPAAAVTILHGAHF 225

RESULT 6
US-08-988-819-5
Sequence 5, Application US/08988819
Patent No. 6054294
GENERAL INFORMATION:
APPLICANT: CHANG, MING-SHI
TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC.
STREET: ONE AMGEN CENTER DRIVE
CITY: THOUSAND OAKS
STATE: CA
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/988,819
FILING DATE: 12-DEC-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/792,019
FILING DATE: 03-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: COOK, ROBERT R.
REGISTRATION NUMBER: 31,602
REFERENCE/DOCKET NUMBER: A-442A
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-988-819-5

Query Match 97.3%; Score 1193; DB 3; Length 225;
Best Local Similarity 96.9%; Pred. No. 1.8e-125;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPFPNEPDPNPRLGASTLPATVDLEWRSNDKLRLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFPNEPDPNPRLGASTLPATVDLEWRSNDKLRLTQNYEAYSHLLCYLRGL 120

Qy 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLGCTPTWTPGPAHSDFLQ 180
Db 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLGCTPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225

RESULT 7
US-09-016-534-5
; Sequence 5, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-016-534-5

Query Match 97.3%; Score 1193; DB 4; Length 225;
Best Local Similarity 96.9%; Pred. No. 1.8e-125;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60
Db 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDILTRYLEHQLRSLAGT 60

Qy 61 YLNYLGPPFPNEPDPNPRLGASTLPATVDLEWRSNDKLRLTQNYEAYSHLLCYLRGL 120
Db 61 YLNYLGPPFPNEPDPNPRLGASTLPATVDLEWRSNDKLRLTQNYEAYSHLLCYLRGL 120

Qy 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLGCTPTWTPGPAHSDFLQ 180
Db 121 NROATAELRSLAHFCTSLQGLSIGVMAALGYPLPQPLGCTPTWTPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225
Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMKQPPAAAVTLHLGAHGF 225

RESULT 8
US-09-106-182-3
; Sequence 3, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yanguu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiostrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-106-182-3

Query Match 10.2%; Score 124.5; DB 3; Length 203;
Best Local Similarity 28.1%; Pred. No. 5.5e-06;
Matches 50; Conservative 24; Mismatches 85; Indels 19; Gaps 5;

Qy 40 IQKTYDILTRYLEHQLRSLAGTYLNYLGPFPNPPRL---GAETLPRAVTVLEVWRS 96
Db 27 IRQTHNLRLLTKYADQLLEEVVQQGEPFGLPGFSPRLPLAGLSGPAPSHAGLPV--- 83

Qy 97 LNDKLRITQNYEAYSHLLCYLRGLNRQAA-----TAEIARRSLAHFCTSLQGLSIGVMA 151
Db 84 ---SERLRQDAALASLPALLDAVRRRQALNPRAPRLRSLEDAARQVRALGAAVETVL 140

Qy 152 AALGY----PLPQPLGCTPTWTPGPAHSDFLQKMDDFWLLKELQTLWRSKDFNRL 205
Db 141 AALGAAARGVPPEPV-ATSLFTSNSAAGVFSKVLGLHVCGLYGEWVSRTEGDLQGL 197

RESULT 9
US-08-233-609-3
; Sequence 3, Application US/08233609

ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894PID2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-443-129-3

```

Query Match          9.7%; Score 118.5; DB 1; Length 203;
Best Local Similarity 28.6%; Pred. No. 2.6e-05;
Matches             51; Conservative 21; Mismatches 88; Indels 17; Gaps 5;

QY      40 IQKTYDLTRYLEHQRLSLAGTLYNLVGLPPNPEDFNPRL--CAETLPRATVDLEYWRS 96
       ||::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       27 IROTHNARLTTKYAQLBEYVOQQGEPFPGSPRPLAGLSGPAPSHAGLPV--- 83

QY      97 LNDKLRLTQNYEAYSHLLCYVLGNLRQA-----TAELRRSLAHFCTSLQGGSIAGW 151
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       84 ---SERLRQDAALSVPALLDAYERQAELNPAPRLRLSLEDAAQRYALGAAVETVL 140

QY     152 AALGYPL--PQPLPGTEPT-WTPGPAHSDFLOKKMDDFWLLKELOTWLWRSAKDPNRL 205
       |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     141 AALCAAARGPGPEPVTVATFTANSTAGISAKVLGHVCGLYGGEWVSRTSGDLAQ 197

```

RESULT 14
US-08-443-952-3
Sequence 3, Application US/08443952
Patent No. 5679545
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,952
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894PID4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881

```

; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-443-952-3

Query Match          9.7%; Score 118.5; DB 1; Length 203;
Best Local Similarity 28.8%; Pred. No. 2.6e-05;
Matches 51; Conservative 21; Mismatches 88; Indels 17; Gaps 5;

QY 40 IOKTYDLTRYLEHQLRSLAGTYLNYLGGPPNEDFNPPRL---GAETLPRATVDLEVWRS 96
DB 27 IROTHNRLRLTKYARQLLEYYVQQQGEPPGLPGFPPLPLAGLSGPAFSAHGLPV--- 83
QY 97 LNDKRLRTQNYEAYSHLLCYLRGLNQAA-----TAEIIRSLAHFCTSLQGLSGIAGVM 151
DB 84 ---SERLRQDAALSVLPALLDVAVRRRQARLNPAPRLRLSLRLEDAARQVRAALGAATVTL 140
QY 152 AALGYPL--PQLPGTGTPT--WTPGPAHSDFLQKMDQDFWLLKELQTLWLRSAKQFNRL 205
DB 141 AALGAAARGPGPEVTVAULTANSTAGTISAKVLGHVCGLYGEWVSRTREGDLGQL 197

RESULT 15
US-08-443-130-3
; Sequence 3, Application US/08443130
; Patent No. 5723585
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffe
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,130
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894P1D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

US-08-443-130-3

```
Query Match          9.7%; Score 118.5; DB 1; Length 203;
Best Local Similarity 28.8%; Pred. No. 2.6e-05;
Matches 51; Conservative 21; Mismatches 88; Indels 17; Gaps 5;

Qy 40 IQTYDLTRYLEHQLRSIAGTYLNYLGGPPFNEPDNPPRL---GAETLPRATVDLEVWRS 96
Db 27 IROTHNLARLLTKYAEQLLEBYVQQGEPFGLPGFSPPLPLAGLSGAPSHAGLPV--- 83

Qy 97 LNDKLRITONVEAYSHLLCYLRGLNRQAA-----TAE LRSLAHFCTSLQGLLSIAGVM 151
Db 84 ---SERLRQDAALSVLPALLDAVRRQAEINPRAPRLRLSLEDAARQVRALGAAVETVL 140

Qy 152 AALGYPL--PQPLFGTEPT-WTPGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRL 205
Db 141 AALGAAARGPGPEPTVATLFTANSTAGIFS AKVLGFHVCGLYGEWVSRTEGDLGQL 197
```

Search completed: January 27, 2003, 15:43:07
Job time : 14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 12532.7 Seconds
(without alignments)
11812.771 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctggagtggtggcgtggc.....cctttgtaagctctctca 5087

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	5087	100.0	5087	6	AR002596	Sequence
2	5087	100.0	5087	6	AX392088	Sequence
3	5087	100.0	5087	9	AF176912	Homo sapi
C	4910.4	96.5	168567	9	AP003419	Homo sapi
C	4908.8	96.5	169144	2	AC005849	Homo sapi
5	4605	90.5	135116	2	AF002437	Homo sapi
6	4605	90.5	135116	2	AF002437	Homo sapi
7	1709.4	33.6	269355	2	AC109138	Mus muscu
8	1463.4	28.8	1736	9	BC012939	Homo sapi
9	1461.4	28.7	1689	9	AF172854	Homo sapi
10	525.4	10.3	881	6	AX205024	Sequence
11	525.4	10.3	881	6	AX205042	Sequence
12	523.4	10.3	797	6	AR002595	Sequence
13	523.4	10.3	797	6	AX392086	Sequence
14	523.4	10.3	797	9	AF176911	Sequence
15	495.4	9.7	680	9	AY049779	Homo sapi
16	492.8	9.7	1692	6	AX205060	Sequence
17	429.6	8.4	819	6	AR002597	Sequence
18	429.6	8.4	819	6	AX392089	Sequence
19	429.6	8.4	819	10	AF176913	Sequence
20	165.2	3.2	283	6	AX202145	Mus muscu
C	148.8	2.9	63347	2	AC110526	Sequence
22	99	1.9	7218	6	I66494	Sequence 14
23	82.8	1.6	179145	9	AC009785	Homo sapi
24	78	1.5	257393	2	AC102125	Homo sapi
25	73.8	1.5	203982	9	AC067854	Homo sapi
C	73.2	1.4	195690	10	AC098726	Mus muscu
26	72	1.4	98179	2	AC117037	Rattus no
C	70.6	1.4	95097	9	AL133232	Human DNA
28	69.4	1.4	196876	2	AC097558	Rattus no
29	68.8	1.4	35848	2	AC011553	Homo sapi
C	68.8	1.4	125020	9	AF429315	Homo sapi
31	68.2	1.3	161984	2	AC115241	Rattus no
C	67.6	1.3	10732	6	E32986	Gene encodi
33	67.6	1.3	163246	2	AC068401	Homo sapi
34	67.6	1.3	174428	9	AC005821	Homo sapi
35	67.2	1.3	125020	9	AF429315	Homo sapi
C	67.2	1.3	125020	9	AF429315	Homo sapi
37	67	1.3	104518	2	AC130185	Ornithorh
C	66.8	1.3	230000	9	AF243527	Homo sapi
38	65.8	1.3	230687	2	AC124433	Mus muscu
C	65.6	1.3	200947	2	AC016961	Homo sapi
C	65	1.3	171624	2	AC095621	Rattus no
41	65	1.3	196283	10	AL672215	Mouse DNA
42	65	1.3	321438	2	AC104989	Homo sapi
C	64.8	1.3	139729	2	AC104100	Mus muscu
C	64.6	1.3	109975	9	AC112497	Homo sapi

ALIGNMENTS

RESULT 1
AR002596
LOCUS
DEFINITION Sequence 3 from patent US 5741772.
ACCESSION AR002596
VERSION AR002596.1 GI:3964150
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 5087)
AUTHORS Chang,M.-s.
TITLE Neurotrophic factor NNT-1
JOURNAL Patent: US 5741772-A 3 21-APR-1998;
FEATURES Location/Qualifiers

AR002596 5087 bp DNA linear PAT 04-DEC-1998

Qy	4261	TTGGCTGCCCACTCTCTTGGATGGGTGTTGCTCCCTTATCCCAAACTCACTCTATACA	4320
Db	4261	TTGGCTGCCCACTCTCTTGGATGGGTGTTGCTCCCTTATCCCAAACTCACTCTATACA	4320
Qy	4321	TCCAAATCAGGAAACAAACATGGTGGCAATTTCTACACAAAAGAGATGAGATTAACAGTG	4380
Db	4321	TCCAAATCAGGAAACAAACATGGTGGCAATTTCTACACAAAAGAGATGAGATTAACAGTG	4380
Qy	4381	CAGGGTGGGGTCTGCAATTTGGAGGTCCCTTATAAACCAGAGAGAAAATACTGAAAGCAC	4440
Db	4381	CAGGGTGGGGTCTGCAATTTGGAGGTCCCTTATAAACCAGAGAGAAAATACTGAAAGCAC	4440
Qy	4441	AGGGCAGGGAACAGACACAGACCCAGGAGTCTCCAAAGCACAGAGTGGGCAACAAA	4500
Db	4441	AGGGCAGGGAACAGACACAGACCCAGGAGTCTCCAAAGCACAGAGTGGGCAACAAA	4500
Qy	4501	ACCCGAGCTGAGCATCAGACACCTTGCTCGAATTGCTTCCAGTATTACGGTGCCTTTC	4560
Db	4501	ACCCGAGCTGAGCATCAGACACCTTGCTCGAATTGCTTCCAGTATTACGGTGCCTTTC	4560
Qy	4561	TCTGCCCTCTTCCAGGGTATCTGTGGTTCAGAGCTGGGAGGCAACCATAGCCAC	4620
Db	4561	TCTGCCCTCTTCCAGGGTATCTGTGGTTCAGAGCTGGGAGGCAACCATAGCCAC	4620
Qy	4621	ACGACAGGATTTCTGAAAGTTTACATGACAGTAGCATTTTGGGGTGTAGGGTGGCAGCT	4680
Db	4621	ACGACAGGATTTCTGAAAGTTTACATGACAGTAGCATTTTGGGGTGTAGGGTGGCAGCT	4680
Qy	4681	CCCCAAGGCTGCCCCCAGCCCCCACTCATGACTCTAAGTGTGTGTAATATAT	4740
Db	4681	CCCCAAGGCTGCCCCCAGCCCCCACTCATGACTCTAAGTGTGTGTAATATAT	4740
Qy	4741	TTATTTATTGGAGATGTTATTTATTAGATGATATTTATTGCGAATTTCTATTCTTGTA	4800
Db	4741	TTATTTATTGGAGATGTTATTTATTAGATGATATTTATTGCGAATTTCTATTCTTGTA	4800
Qy	4801	TTAACAATAAATGTTTCCCAAGACTAGTCTCTTTGGCCAGGCTCACCCCTCTCTGG	4860
Db	4801	TTAACAATAAATGTTTCCCAAGACTAGTCTCTTTGGCCAGGCTCACCCCTCTCTGG	4860
Qy	4861	TGCTCATCAGACTCTTGCCACCCCTGGCTCCACTCCCTGCTTCCCTTGGTGGAGTGC	4920
Db	4861	TGCTCATCAGACTCTTGCCACCCCTGGCTCCACTCCCTGCTTCCCTTGGTGGAGTGC	4920
Qy	4921	ACAGAGCTCTGGGAAGAGGCTTCTCTCCCGCACTGGGGCGATGGGGCAGCTCAGA	4980
Db	4921	ACAGAGCTCTGGGAAGAGGCTTCTCTCCCGCACTGGGGCGATGGGGCAGCTCAGA	4980
Qy	4981	CTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCACACAGCTTC	5040
Db	4981	CTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCACACAGCTTC	5040
Qy	5041	TGTCACCCAGGTTTCCCTCACCCCACTTTGCTTAAGTCTTCTCA 5087	
Db	5041	TGTCACCCAGGTTTCCCTCACCCCACTTTGCTTAAGTCTTCTCA 5087	

RESULT 2	AX392088	Sequence 3 from Patent WO0215977.	5087 bp	DNA	linear	PAT 23-MAR-2002
LOCUS	AX392088					
DEFINITION	AX392088					
ACCESSION	AX392088.1	GI:19700576				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
	human.					
	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE						
AUTHORS	Senaldi, G.					
TITLE	Methods and compositions for treating ige-related disease using nn					
	t-i inhibitors					

JOURNAL Patent: WO 0215977-A 3 28-FEB-2002;

Amgen Inc. (US)

FEATURES Location/Qualifiers

source

1..5087

/organism="Homo sapiens"

/db_xref="taxon:9606"

137..138

misc_feature

/note="INTERVENING UNSEQUENCED REGION OF >1KB"

BASE COUNT 992 a 1746 c 1191 g 1158 t

ORIGIN

Query Match 100.0%; Score 5087; DB 6; Length 5087;

Best Local Similarity 100.0%; Pred. NO. 0;

Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACCTGGAGTGGGCTGGGATGGGATTTAAAGCTTCGCCGAGCGCGGCTCGCC 60

Db 1 AACCTGGAGTGGGCTGGGATGGGATTTAAAGCTTCGCCGAGCGCGGCTCGCC 60

Qy 61 CTCCTCCAGCTCCGCGAGAGGAGCGCACCGGCGGCGCCAGCCGCCA 120

Db 61 CTCCTCCAGCTCCGCGAGAGGAGCGCACCGGCGGCGCCAGCCGCCA 120

Qy 121 TGGACCTCCGAGAGGTTGAAAACCCAACTAGCCCTCTTCAATACATGACAGCAG 180

Db 121 TGGACCTCCGAGAGGTTGAAAACCCAACTAGCCCTCTTCAATACATGACAGCAG 180

Qy 181 CGCCCTCATCTGATACCTAAACCGCAAGTCAACGCTCCAACTCACCTCTGCTGCC 240

Db 181 CGCCCTCATCTGATACCTAAACCGCAAGTCAACGCTCCAACTCACCTCTGCTGCC 240

Qy 241 CAGACCTCACACATCTTGTGGACTCAAACTCAACCGCACTAAATCAACCAATCCCA 300

Db 241 CAGACCTCACACATCTTGTGGACTCAAACTCAACCGCACTAAATCAACCAATCCCA 300

Qy 301 AGCTAAACTAATCTGAACTTTTAAAGTAAACCCAGTCTTAACTAACCTAGCCCAAT 360

Db 301 AGCTAAACTAATCTGAACTTTTAAAGTAAACCCAGTCTTAACTAACCTAGCCCAAT 360

Qy 361 GCCAATTATCTACCTAGCCAAACCCCTAACTGCTTTCGCACTCCAAAGTGTCCACTG 420

Db 361 GCCAATTATCTACCTAGCCAAACCCCTAACTGCTTTCGCACTCCAAAGTGTCCACTG 420

Qy 421 AATCTCATCTTGTGCTCACTGAAAATCCAGAAAAGCATATTTCCCACTGCCACAT 480

Db 421 AATCTCATCTTGTGCTCACTGAAAATCCAGAAAAGCATATTTCCCACTGCCACAT 480

Qy 481 CCCTCTTACAGCACCCCAACCTGCTGCTCTGCACTCTGCTATCTCTGGATGTCCTAACT 540

Db 481 CCCTCTTACAGCACCCCAACCTGCTGCTCTGCACTCTGCTATCTCTGGATGTCCTAACT 540

Qy 541 CTGCACTGCGCATCAGCCAAACAGCCGACTCGTCAAAATGACCTCTCTCTCTCTGCTCC 600

Db 541 CTGCACTGCGCATCAGCCAAACAGCCGACTCGTCAAAATGACCTCTCTCTCTCTGCTCC 600

Qy 601 CCACCTTTCAGCTGATGAAAAGGCTCATTTGAAGTCCAACTTTCCCACTTAAACAC 660

Db 601 CCACCTTTCAGCTGATGAAAAGGCTCATTTGAAGTCCAACTTTCCCACTTAAACAC 660

Qy 661 AAGACGGGGTGAACCTCCACACTGCAACCGTTCCTGAGAGTGAAGTAAATCTCCTT 720

Db 661 AAGACGGGGTGAACCTCCACACTGCAACCGTTCCTGAGAGTGAAGTAAATCTCCTT 720

Qy 721 CAATCTAACCCCAACCTTACACTTCCCACTCAGGAATCATCTCTAGATATATACCCAAA 780

Db 721 CAATCTAACCCCAACCTTACACTTCCCACTCAGGAATCATCTCTAGATATATACCCAAA 780

Qy 781 ACTAAGCCCATTAAGCAGCGCCGACCTAGTGTCTAACCTTACCTTCTCTCTCTATGG 840

Db 781 ACTAAGCCCATTAAGCAGCGCCGACCTAGTGTCTAACCTTACCTTCTCTCTCTATGG 840

Qy 841 GTGAGTCTGTTCTGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900

Db 841 GTGAGTCTGTTCTGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900

Qy 3061 CCAAGTTTGGCCACCCAGCCCAATCAGCAGCTCTTCTCTGAGCGTCTTGCGCTCTCCTC 3120
Db CCAAGTTTGGCCACCCAGCCCAATCAGCAGCTCTTCTCTGAGCGTCTTGCGCTCTCCTC 3120
Qy 3121 CTCCTCTCTTTTCTAGCCCTCTCCATTGGAGAGCTCACACCGCCACTGCTTCAACTGTC 3180
Db CTCCTCTCTTTTCTAGCCCTCTCCATTGGAGAGCTCACACCGCCACTGCTTCAACTGTC 3180
Qy 3181 ACCTGCATCAAAATGATATCTTATTGGAAAACTCAGGGAGGCCATGAACAAAGAGCC 3240
Db ACCTGCATCAAAATGATATCTTATTGGAAAACTCAGGGAGGCCATGAACAAAGAGCC 3240
Qy 3241 TAGCATGGAGACAGGCCCACTGTACGGGACACAAAAATAGAAACTTTGGGAGCAGGTA 3300
Db TAGCATGGAGACAGGCCCACTGTACGGGACACAAAAATAGAAACTTTGGGAGCAGGTA 3300
Qy 3301 TCTCCTTGGTGTGAGCAGCGGCTCTGCGCTCTCTTCCCTCATCAGCTCTCTCTTTTC 3360
Db TCTCCTTGGTGTGAGCAGCGGCTCTGCGCTCTCTTCCCTCATCAGCTCTCTCTTTTC 3360
Qy 3361 ACAGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGG 3420
Db ACAGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGG 3420
Qy 3421 GGCAGAGACTCTGCCAGGGCCACTGTGATTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480
Db GGCAGAGACTCTGCCAGGGCCACTGTGATTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480
Qy 3481 ACTGGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTACTTTGGTGGCCCT 3540
Db ACTGGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTACTTTGGTGGCCCT 3540
Qy 3541 CAACCGTCAAGGCTGCCACTGTGAGCTGGCGCCAGCTTCAACCCCTCCCGCCCTGGG 3600
Db CAACCGTCAAGGCTGCCACTGTGAGCTGGCGCCAGCTTCAACCCCTCCCGCCCTGGG 3600
Qy 3601 CAGGGCTGTGGGAGCAATTCGGGCGTCAATGGAGCTCTGGGCTACCCACTGCCCCA 3660
Db CAGGGCTGTGGGAGCAATTCGGGCGTCAATGGAGCTCTGGGCTACCCACTGCCCCA 3660
Qy 3661 GCGGCTGCTGGGACTGAAACCACTTGGACTCTCTGGCCCTGCCACAGTACTTCTCCA 3720
Db GCGGCTGCTGGGACTGAAACCACTTGGACTCTCTGGCCCTGCCACAGTACTTCTCCA 3720
Qy 3721 GAAGATGACGACTTCTGGCTGTGAAGGAGCTGCAGACTGTGGCGCTCGGCCAA 3780
Db GAAGATGACGACTTCTGGCTGTGAAGGAGCTGCAGACTGTGGCGCTCGGCCAA 3780
Qy 3781 GGAATTTCAACCGGCTCAAGAGAAGATGAGAGCTTCCAGAGCTGCAGTCAACCTGACCT 3840
Db GGAATTTCAACCGGCTCAAGAGAAGATGAGAGCTTCCAGAGCTGCAGTCAACCTGACCT 3840
Qy 3841 GGGGCTCATGGCTCTGACTCTGACTCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900
Db GGGGCTCATGGCTCTGACTCTGACTCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900
Qy 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCAACACCTGTGTAGCCAGGAGACAGAGCTG 3960
Db CCCACTTTGTAGAGCCAGCCCTGTATGCAACACCTGTGTAGCCAGGAGACAGAGCTG 3960
Qy 3961 TGAGCTCTGGCCCTTTCTTGACCGGCTGGGCGTGTGATGCGATCAGCCCTCTCTCCTC 4020
Db TGAGCTCTGGCCCTTTCTTGACCGGCTGGGCGTGTGATGCGATCAGCCCTCTCTCCTC 4020
Qy 4021 CCCACTCCCAAGGCTACCGAGCTGGGAGGAGTACAGTAGGCGCTCTCTCTCTCTCTCT 4080
Db CCCACTCCCAAGGCTACCGAGCTGGGAGGAGTACAGTAGGCGCTCTCTCTCTCTCTCT 4080
Qy 4081 TTTTACAGGAAGTCAATGCTCAGAGGAGTGTGAAGTGGTTTCAAGTGTGTGAGAGGCGCT 4140
Db TTTTACAGGAAGTCAATGCTCAGAGGAGTGTGAAGTGGTTTCAAGTGTGTGAGAGGCGCT 4140
Qy 4141 CATGGCTCTCTGCTTCTTGCTTACCACTTGGGCGAGTGGCCACCCAGCCCTCAGGTGGCA 4200

Db 4141 CATGGCTCTCTGCTTCTTGCTTACCACTTGGCCAGTGGCCACCCAGCCCTCAGGTGGCA 4200
Qy 4201 CATCTGAGGAGGAGGCTTGAAGGGCCACACCAACATGCTTCTTCTGGGTGAAGCCCT 4260
Db CATCTGAGGAGGAGGCTTGAAGGGCCACACCAACATGCTTCTTCTGGGTGAAGCCCT 4260
Qy 4261 TTGGCTGCCCACTCTCTCTTGGATGGGTGTGCTTATCCCAAAATCACTCTATACA 4320
Db TTGGCTGCCCACTCTCTCTTGGATGGGTGTGCTTATCCCAAAATCACTCTATACA 4320
Qy 4321 TCAATTCAGGAACAAACATATGCTGGCAATTTCTACA CAAAAGAGATGAGATTACAGTG 4380
Db TCAATTCAGGAACAAACATATGCTGGCAATTTCTACA CAAAAGAGATGAGATTACAGTG 4380
Qy 4381 CAGGTTGGGCTCTGCAATGGAGGTGCCCTATAACACAGAGAGAAAAATCTGAAAGCAC 4440
Db CAGGTTGGGCTCTGCAATGGAGGTGCCCTATAACACAGAGAGAAAAATCTGAAAGCAC 4440
Qy 4441 AGGGCAGGGACAGACAGACAGACAGACAGACAGACAGACAGACAGAGTGGCAACAA 4500
Db AGGGCAGGGACAGACAGACAGACAGACAGACAGACAGACAGAGTGGCAACAA 4500
Qy 4501 ACCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGTCTTCCAGTATTAGGTGCTCTTC 4560
Db ACCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGTCTTCCAGTATTAGGTGCTCTTC 4560
Qy 4561 TCTGCCCTCTTCCAGGTAATCTGTGGTTGCCAGGCTGGGAGGGCAACCATAGCCAC 4620
Db TCTGCCCTCTTCCAGGTAATCTGTGGTTGCCAGGCTGGGAGGGCAACCATAGCCAC 4620
Qy 4621 ACCACAGATTTCTGAAAGTTTACAATGCAATGAGTACTTTTGGGGTGTAGGTGGCAGCT 4680
Db ACCACAGATTTCTGAAAGTTTACAATGCAATGAGTACTTTTGGGGTGTAGGTGGCAGCT 4680
Qy 4681 CCCCAAGGCTTGGCCCGCCAGCCCAACCATGACTCTAAAGTGTGTGTTAATAT 4740
Db CCCCAAGGCTTGGCCCGCCAGCCCAACCATGACTCTAAAGTGTGTGTTAATAT 4740
Qy 4741 TTAATTTATTTGGAGATGTTATTTATAGATGATATTTATGAGAAATTTCTATTCTTGA 4800
Db TTAATTTATTTGGAGATGTTATTTATAGATGATATTTATGAGAAATTTCTATTCTTGA 4800
Qy 4801 TTAACAAATAAATGTTTGGCCCAAGAACTTAGTCTCTTTTGGCCAGCCTCA CCCCCTCTGG 4860
Db TTAACAAATAAATGTTTGGCCCAAGAACTTAGTCTCTTTTGGCCAGCCTCA CCCCCTCTGG 4860
Qy 4861 TGCTCATCAGACTTTCGACACCCCTGGCTCCCACTCCCTGCTTGGCTCTGGTGGAGCTGC 4920
Db TGCTCATCAGACTTTCGACACCCCTGGCTCCCACTCCCTGCTTGGCTCTGGTGGAGCTGC 4920
Qy 4921 ACAGAGCTCTGGGAAGAGGCCCTCTTCTCCCGCAGCTGGGCGATGGCGCACCTCAGA 4980
Db ACAGAGCTCTGGGAAGAGGCCCTCTTCTCCCGCAGCTGGGCGATGGCGCACCTCAGA 4980
Qy 4981 CTTTACCCACTGCTCTGCGCACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040
Db CTTTACCCACTGCTCTGCGCACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040
Qy 5041 TGTCACCCAGGTTTCTCTCACCCAGCTTGTGTAAGTCTTCCTCA 5087
Db TGTCACCCAGGTTTCTCTCACCCAGCTTGTGTAAGTCTTCCTCA 5087

RESULT 3

AF176912

LOCUS

DEFINITION

Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene,

complete cds.

ACCESSION

AF176912

VERSION

AF176912.1 GI:6007642

KEYWORDS

SOURCE

5087 bp DNA linear

PRI 04-OCT-1999

Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene,

complete cds.

AF176912

AF176912.1 GI:6007642

Homo sapiens.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 5087) Senaldi, G., Varnum, B.C., Sarmiento, U., Lile, J., Starnes, C., Lile, J., Scully, S., Guo, J., Elliott, G., McNinch, J., Shaklee, C.L., Freeman, D., Manu, F., Simonet, W.S., Boone, T. and Chang, M.-S.
TITLE	Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the IL-6 family
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)
MEDLINE	99432254
PUBMED	10500198
REFERENCE	2 (bases 1 to 5087)
AUTHORS	Senaldi, G., Varnum, B., Sarmiento, U., Lile, J., Starnes, C., Scully, S., Guo, J., Elliott, G., McNinch, J., Freeman, D., Shaklee, C., Manu, F., Simonet, S., Boone, T. and Chang, M.-S.
TITLE	Direct Submission
JOURNAL	Submitted (11-AUG-1999) Angen, Inc., One Angen Center Drive, Thousand Oaks, CA 91320, USA
FEATURES	Location/Qualifiers
source	1. 5087 /organism="Homo sapiens" /db_xref="taxon:9606" join(<120..135,1368..1534,3365..1534,3365..>.3859) /product="neurotrophin-1/B-cell stimulating factor-3" join(120..135,1368..1534,3365..1534,3365..3859) /codon_start=1 /product="neurotrophin-1/B-cell stimulating factor-3" /protein_id="AAFO0992.1" /db_xref="GI:6007643" /translation="MDLRAGDSWGMLACLTVLWHLPAPVAPALNRTGDPGPGSIQKYDLTRYLHQURLSLAGTIVYLPFNEPDPFNPRLGAETLPRAIVDELYVRSINDKLRLTONYEAYSHLLCVLRGLNRQAAETLRRSLAHFCTSLQGLIGSIAGVMAALGYPLQPLPCTEPTWTPGPAHSDFLQKMDDFLLKELQTLWLRSAKDFNRLKKQWQPPAAAVTLHLGANGF"
BASE COUNT	992 a 1746 c 1191 g 1158 t
ORIGIN	
Query Match	100.0%; Score 5087; DB 9; Length 5087;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 5087; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 AACCTGGAGTGGGCGTGGCGGATGGGATTATTAAGCTTCGCGGAGCGCGCTCGCC 60
Db	1 AACCTGGAGTGGGCGTGGCGGATGGGATTATTAAGCTTCGCGGAGCGCGCTCGCC 60
Qy	61 CTCCCACTCCGAGCCTCGGAGAGAGCGGCACCGGCGGCGGCGCCAGCCCCA 120
Db	61 CTCCCACTCCGAGCCTCGGAGAGAGCGGCACCGGCGGCGGCGCCAGCCCCA 120
Qy	121 TGGACCTCCGAGAGTGTGAAACCCAACTAGCCCTGCTCTTCATACATGACAACGAG 180
Db	121 TGGACCTCCGAGAGTGTGAAACCCAACTAGCCCTGCTCTTCATACATGACAACGAG 180
Qy	181 CGCCCCATCTGATACCTAAACCGACCAAGTCACAGCCCTCCAACTACCCCTGTCCTGCC 240
Db	181 CGCCCCATCTGATACCTAAACCGACCAAGTCACAGCCCTCCAACTACCCCTGTCCTGCC 240
Qy	241 CAGACCTCACCATCTCTTGGACTCAAACTCAACCGCACTAAATCAACCAATCCCA 300
Db	241 CAGACCTCACCATCTCTTGGACTCAAACTCAACCGCACTAAATCAACCAATCCCA 300
Qy	301 AGTCTAACTAATCTGAAACTTTTAAAGTACCAGTCCCTTAAACCTTAACCTAGCCCAAT 360
Db	301 AGTCTAATCTAATCTGAAACTTTTAAAGTACCAGTCCCTTAAACCTTAACCTAGCCCAAT 360
Qy	361 GCCAATTATATCTACCCCTAGCCAAACCCCTAACTGCCTTTGCGAGTCCAAAGTGTCCACTG 420
Db	361 GCCAATTATATCTACCCCTAGCCAAACCCCTAACTGCCTTTGCGAGTCCAAAGTGTCCACTG 420
Qy	421 AATCCTCACCTTGTCCTCACTGAAAATCCAGAAAAAGCATATTTTCCCACTGCCCAAT 480
Db	421 AATCCTCACCTTGTCCTCACTGAAAATCCAGAAAAAGCATATTTTCCCACTGCCCAAT 480

QY 1561 AGTTGGGAGGAGTGGAGTTGGGAAAGACAGTCCTAACCGTGGAGGTTCTGGTAAA 1620
DB 1561 AGTTGGGAGGAGTGGAGTTGGGAAAGACAGTCCTAACCGTGGAGGTTCTGGTAAA 1620
QY 1621 TGATGGGTGAGGAGGGGCTTTTGGCTCCCAACAGTCCCTGTCTGTCTATCTCCTG 1680
DB 1621 TGATGGGTGAGGAGGGGCTTTTGGCTCCCAACAGTCCCTGTCTGTCTATCTCCTG 1680
QY 1681 CCGTTCCCTCTTAGGTGGCCCCCCCCACATTCCTCCATCCCTGGCCCCAGGACTAGGCAITG 1740
DB 1681 CCGTTCCCTCTTAGGTGGCCCCCCCCACATTCCTCCATCCCTGGCCCCAGGACTAGGCAITG 1740
QY 1741 GGCAGGCTCGCACCGGCTTGGCCCATTTGCCCATTTGCTGCTGCGAGCCAGCGCCCGCC 1800
DB 1741 GGCAGGCTCGCACCGGCTTGGCCCATTTGCCCATTTGCTGCTGCGAGCCAGCGCCCGCC 1800
QY 1801 TCCCTCTGGGGCCGGGAAGTCTCTCTGTGTTTACACCGTGTGTGTCTCTTGGCGG 1860
DB 1801 TCCCTCTGGGGCCGGGAAGTCTCTCTGTGTTTACACCGTGTGTGTCTCTTGGCGG 1860
QY 1861 GCGGGGTTGGTGGGACAGAGGGGCCACCTCCATGCTGCGTTCCAGTCCGCTC 1920
DB 1861 GCGGGGTTGGTGGGACAGAGGGGCCACCTCCATGCTGCGTTCCAGTCCGCTC 1920
QY 1921 TGCCCCCAGACCTGGGGCCCTGCTGCTGTGACCCAGGGGCTTCCCTTCCGTCGCTCT 1980
DB 1921 TGCCCCCAGACCTGGGGCCCTGCTGCTGTGACCCAGGGGCTTCCCTTCCGTCGCTCT 1980
QY 1981 CCGATCTAGTGGGCTCTAGGGGGTCTATGGGGGATGGGGAGGGGACTGTAGGGAAACCCAGG 2040
DB 1981 CCGATCTAGTGGGCTCTAGGGGGTCTATGGGGGATGGGGAGGGGACTGTAGGGAAACCCAGG 2040
QY 2041 CAGTAGTGCAGGGGTTTAGGGTGTGATGAGGTTATGCTGTAAGGATTTTGGGGGTG 2100
DB 2041 CAGTAGTGCAGGGGTTTAGGGTGTGATGAGGTTATGCTGTAAGGATTTTGGGGGTG 2100
QY 2101 TCCAGAGTGTTCAGAGAGCCAGGAGAGAGGAGGAGGTTGGAGAGCCGAGGACCC 2160
DB 2101 TCCAGAGTGTTCAGAGAGCCAGGAGAGAGGAGGAGGTTGGAGAGCCGAGGACCC 2160
QY 2161 ATGGGGAACCGGCCCTCTTCCGTTGTTCTTCCATGCCAGACCTACTCTGGAG 2220
DB 2161 ATGGGGAACCGGCCCTCTTCCGTTGTTCTTCCATGCCAGACCTACTCTGGAG 2220
QY 2221 CCAGGGAAGAAAGGAGAGAGTGCGGGGAGTGGCTCCAGCCCCAGGATACACCG 2280
DB 2221 CCAGGGAAGAAAGGAGAGAGTGCGGGGAGTGGCTCCAGCCCCAGGATACACCG 2280
QY 2281 AGGAAATTAGTTTGTCTGTGCTGTGTCAGCGTGTGAACCTCCCTGGGCCCTTGCCTA 2340
DB 2281 AGGAAATTAGTTTGTCTGTGCTGTGTCAGCGTGTGAACCTCCCTGGGCCCTTGCCTA 2340
QY 2341 TCCAGGCTCTCCCTTGTCTCTTCTTCTTCCAGATTATACATCTCCCTCATCCCTT 2400
DB 2341 TCCAGGCTCTCCCTTGTCTCTTCTTCTTCCAGATTATACATCTCCCTCATCCCTT 2400
QY 2401 TCCCTGGGCCAGCGCTCCCGAGGGTTGAAAGGGCTCTGCCCTCTTCCCTATACC 2460
DB 2401 TCCCTGGGCCAGCGCTCCCGAGGGTTGAAAGGGCTCTGCCCTCTTCCCTATACC 2460
QY 2461 ATGCTGTCTTCCATAGCCTTCTCTCTGTCTACTATGAGACTGCCTCCATTTCTTCTT 2520
DB 2461 ATGCTGTCTTCCATAGCCTTCTCTCTGTCTACTATGAGACTGCCTCCATTTCTTCTT 2520
QY 2521 CTGCAACCTTCTCTATAGCTGAACCTTCTTCTGGAGTGTAGTGAACCTCTC 2580
DB 2521 CTGCAACCTTCTCTATAGCTGAACCTTCTTCTGGAGTGTAGTGAACCTCTC 2580
QY 2581 TCCCGAGCCCTCAGCTGTGGGCTGTGGTGTGTGTCAGCGGCAATGGGGCTCTGGTTCCA 2640
DB 2581 TCCCGAGCCCTCAGCTGTGGGCTGTGGTGTGTGTCAGCGGCAATGGGGCTCTGGTTCCA 2640
QY 2641 ATGGGCCCATCTCATCTCTCTTGTCTTGTGTCAGAAAAACCTTGTCTCACTCCACTG 2700

DB 2641 ATGGGCCCATCTCATCTCTCTTGTGTCAGAAAAACCTTGTCTCACTCCACTG 2700
QY 2701 CCGTCTCTAGTTCCCGACCCCTTTTCTCTCTCTGCTTCCCTGCAAAATTTCTCAAGGA 2760
DB 2701 CCGTCTCTAGTTCCCGACCCCTTTTCTCTCTGCTTCCCTGCAAAATTTCTCAAGGA 2760
QY 2761 GTGGTCTACACCTCTGCTCCACTTCTCTCCACCACTCACTTCTTAAACCCCTGCAA 2820
DB 2761 GTGGTCTACACCTCTGCTCCACTTCTCTCCACCACTCACTTCTTAAACCCCTGCAA 2820
QY 2821 TCTGGCTTCAGGCCCCAGCAATGGTTCTCTCAAGTCTCAGGCACTCTTGGCCAG 2880
DB 2821 TCTGGCTTCAGGCCCCAGCAATGGTTCTCTCAAGTCTCAGGCACTCTTGGCCAG 2880
QY 2881 CCGACAGTCTTTTGAAGGCTCAATCTCTTCTGCTGTCTTGTGAGCCACACTGCTGAG 2940
DB 2881 CCGACAGTCTTTTGAAGGCTCAATCTCTTCTGCTGTCTTGTGAGCCACACTGCTGAG 2940
QY 2941 CGTCTGCTCTCTCGAACTCTCTTCTTGGTCTCTGCACTCTCTGGGCCACTTCTA 3000
DB 2941 CGTCTGCTCTCTCGAACTCTCTTCTTGGTCTCTGCACTCTCTGGGCCACTTCTA 3000
QY 3001 CCGTCTGCTCTCTCGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3060
DB 3001 CCGTCTGCTCTCTCGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3060
QY 3061 CCAAGGTTTCCGACCCAGCAATCAGACGCTCTCTCTGAGCGTCTTGTGCTCTCTCT 3120
DB 3061 CCAAGGTTTCCGACCCAGCAATCAGACGCTCTCTCTGAGCGTCTTGTGCTCTCTCT 3120
QY 3121 CTCTCTCTTTTCTAGCGCTCTCCATTTGAGAGTCTACCAACCGCACTGCTTCAATGTC 3180
DB 3121 CTCTCTCTTTTCTAGCGCTCTCCATTTGAGAGTCTACCAACCGCACTGCTTCAATGTC 3180
QY 3181 ACTGCAATCAATGATATCTTATTTGAAAACTCAGGAGGCCATGAACAAAGGCC 3240
DB 3181 ACTGCAATCAATGATATCTTATTTGAAAACTCAGGAGGCCATGAACAAAGGCC 3240
QY 3241 TAGCATGGAGACAGGGCCAGTGTCTAGGGGACAAAAAATAGAACTTTGGAGCAGGTA 3300
DB 3241 TAGCATGGAGACAGGGCCAGTGTCTAGGGGACAAAAAATAGAACTTTGGAGCAGGTA 3300
QY 3301 TCTCTTGGTGGTGGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360
DB 3301 TCTCTTGGTGGTGGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360
QY 3361 ACAGTGAACTACTTGGGCCCTTTTCAACAGGCGACTTCAACCTCCCGCTGGG 3420
DB 3361 ACAGTGAACTACTTGGGCCCTTTTCAACAGGCGACTTCAACCTCCCGCTGGG 3420
QY 3421 GGCAAGACTCTGCGCAGGGCCACTTGTGACTTGGAGGTGGGCAAGCCTCAATGACAA 3480
DB 3421 GGCAAGACTCTGCGCAGGGCCACTTGTGACTTGGAGGTGGGCAAGCCTCAATGACAA 3480
QY 3481 ACTCGGCTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTCTTACTTCTGGCT 3540
DB 3481 ACTCGGCTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTCTTACTTCTGGCT 3540
QY 3541 CAACCGTCAAGCTGCTGCTGCGCGGAGCTGCGGCGGAGCTTCTGTCACAGGCT 3600
DB 3541 CAACCGTCAAGCTGCTGCTGCGCGGAGCTGCGGCGGAGCTTCTGTCACAGGCT 3600
QY 3601 CCAGGCTCTGCTGGGAGCTTGGGGGTCTAGTGGAGTCTGGGCTACCCACTGCCCCA 3660
DB 3601 CCAGGCTCTGCTGGGAGCTTGGGGGTCTAGTGGAGTCTGGGCTACCCACTGCCCCA 3660
QY 3661 GCCGCTGCTGGGACTGAACCCACTTGAACCTCTGGGCGCTGCCACAGTCTTCTCCA 3720
DB 3661 GCCGCTGCTGGGACTGAACCCACTTGAACCTCTGGGCGCTGCCACAGTCTTCTCCA 3720
QY 3721 GAAGATGACGACTTCTGGCTGCTGAGAGGCTGAGACCTTGGCTGTGGCTCGGCCAA 3780

Db	49146	ATCTGAACACTTTTAAAGTAACCCAGTCCTTTAAACCTTAACCTAGGCCCAATGCCAATTATAT	49081
Qy	372	CTACCCCTAGGCCAAACCCCTAACTGCCTTTTGGCAGTCCAAAGTGTCCACTGAATCCTCACCT	431
Db	49086	CTACCCCTAGCCAAACCCCTAACTGCCTTTTGGCAGTCCAAAGTGTCCACTGAATCCTCACCT	49027
Qy	432	TGGTCTCTCACTGAAATCCCHAGAAAGGATATTTCCGCACATGCCCAATCCCTCTCTTACA	491
Db	49026	TGGTCTCTCACTTAAATCCCGAAGGATATTTCCGCACATGCCCAATCCCTCTCTTACA	48967
Qy	492	GCACCCAAACCTGCCTCTGGACTCCTGGTATCTCTGGATGTCCAAACTCTGCAGTGCCTA	551
Db	48966	GCACCCAAACCTGCCTCTGGACTCCTGGTATCTCTGGATGTCCAAACTCTGCAGTGCCTA	48907
Qy	552	TCAGCCAAACAGCCCGACTCGTCAAATGCACCTCTCTCCCTTCTGTCCCCACACCTTTGCA	611
Db	48906	TCAGCCAAACAGCCCGACTCGTCAAATGCACCTCTCTCCCTTCTGTCCCCACACCTTTGCA	48847
Qy	612	GGCTGATGGAAAGGCGCTCATTTGAAGTCCAACTTTTCCACCTTAACACCAAGAACGGGGT	671
Db	48846	GGCTGATGGAAAGGCGCTCATTTGAAGTCCAACTTTTCCACCTTAACACCAAGAACGGGGT	48787
Qy	672	GAACTCTCCACACTGCCACCGTTCCCTGAGAGTGAGCACTAAATCTCTCTCAATCTTAACCC	731
Db	48786	GAACTCTCCACACTGCCACCGTTCCCTGAGAGTGAGCACTAAATCTCTCTCAATCTTAACCC	48727
Qy	732	CACCTTACACTTCCACACTCAGGAATCACATCCTAGAATATACCCAAAATTAAGCCCTCA	791
Db	48726	CACCTTACACTTCCACACTCAGGAATCACATCCTAGAATATACCCAAAATTAAGCCCTCA	48667
Qy	792	TAAAGCAGCCCGACCCCTAGTGGTCTAAACCTTATACCTTGTCTCTATGGGTGAGTCTGTT	851
Db	48666	TAAAGCAGCCCGACCCCTAGTGGTCTAAACCTTATACCTTGTCTCTATGGGTGAGTCTGTT	48607
Qy	852	CTTGGGGCCGCGCTCTCTCTGCTTCTCTCCCTTAGAGCTGACTGTCTCAGCTGCCAGC	911
Db	48606	CTTGGGGCCGCGCTCTCTCTGCTTCTCTCCCTTAGAGCTGACTGTCTCAGCTGCCAGC	48547
Qy	912	TCTGACATGTGCTGTCTCCACCCCTCTGACTCCCTCAAGCTGCAGTGGGACTCGAAGAC	971
Db	48546	TCTGACATGTGCTGTCTCCACCCCTCTGACTCCCTCAAGCTGCAGTGGGACTCGAAGAC	48488
Qy	972	TGGCAGGAAGCTTAGGGTACAACCTGSAACAAGGCAGGTGCACTTCCCTAGGCGCTG	1031
Db	48487	TGGCAGGAAGCTTAGGGTACAACCTGSAACAAGGCAGGTGCACTTCCCTAGGCGCTG	48428
Qy	1032	GCCCCGTCCCTCATGTATACACATATATATGTTGGGACACACACAGTGGGCACATGCC	1091
Db	48427	GCCCCGTCCCTCATGTATACACATATATATGTTGGGACACACACAGTGGGCACATGCC	48368
Qy	1092	AAAGACTCTCTCAGCTGACACACAGATCCATTCTCAAGTATCTACTGATAGACACTCATG	1151
Db	48367	AAAGACTCTCTCAGCTGACACACAGATCCATTCTCAAGTATCTACTGATAGACACTCATG	48308
Qy	1152	CGTGCCAAAGTCTCATCTCTCAAACATACATGCCCTCTCTTTCTCTCCCGCTTTGCCAGG	1211
Db	48307	CGTGCCAAAGTCTCATCTCTCAAACATACATGCCCTCTCTTTCTCTCCCGCTTTGCCAGG	48248
Qy	1212	AGTGTGTTCCCTCTCTCATTCCTCTGCTGCCATCTGGGTGCCACCTCATCCCCCACC	1271
Db	48247	AGTGTGTTCCCTCTCTCATTCCTCTGCTGCCATCTGGGTGCCACCTCATCCCCCACC	48188
Qy	1272	CCAGCCCAAGTGGGGACAGACCTGAGGGGCTGCCAGCTGCTTCCCGCTGTGGGCCCG	1331
Db	48187	CCAGCCCAAGTGGGGACAGACCTGAGGGGCTGCCAGCTGCTTCCCGCTGTGGGCCCG	48128
Qy	1332	GGCCCGCGCTCATGCTTCTCGTCCATCTCTGCCACAGGGGACTCGTGGGGGATGTTAGCGT	1391
Db	48127	GGCCCGCGCTCATGCTTCTCGTCCATCTCTGCCACAGGGGACTCGTGGGGGATGTTAGCGT	48068
Qy	1392	GCCTGTGACCGGTGCTGTGGACCTCTCTGTGAGTGCAGCTCTCAATCGCACAGGGGACC	1451
Db	48067	GCCTGTGACCGGTGCTGTGGACCTCTCTGTGAGTGCAGCTCTCAATCGCACAGGGGACC	48008

Db 144122 AATGATATCCTTATTTGAAAAAATCAGGGAGGCCATGAAACAAAGAGCCCTAGCATGGAGA 144063
Qy 3252 CAGGCGCAGTGTGAGGGGACACAAAAAATAGAAATCTTTGGGAGCAGGTATCTCTTGGTG 3311
Db 144062 CAGGCGCAGTGTGAGGGGACACAAAAAATAGAAATCTTTGGGAGCAGGTATCTCTTGGTG 144003
Qy 3312 GTGAGCAGCGGTCTGCGCTCTCTCTCTCCCATCACCCTCTCTCTTTTACAGCTGAAC 3371
Db 144002 GTGAGCAGCGGTCTGCGCTCTCTCTCTCCCATCACCCTCTCTCTTTTACAGCTGAAC 143943
Qy 3372 ACTTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGCTC 3431
Db 143942 ACTTGGGCCCCCTTTCAACAGCAGCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGCTC 143883
Qy 3432 TGCCCAAGGCGCACTGTGTGACTTGGAGGTGTGGGAGGCTCAATGACAAACTGGGGCTGA 3491
Db 143882 TGCCCAAGGCGCACTGTGTGACTTGGAGGTGTGGGAGGCTCAATGACAAACTGGGGCTGA 143823
Qy 3492 CCAGAACTACGAGGCGCTACAGCCACCTCTGTGTACTTGTGCGTGGGCTCAACCGCTCAGG 3551
Db 143822 CCAGAACTACGAGGCGCTACAGCCACCTCTGTGTACTTGTGCGTGGGCTCAACCGCTCAGG 143763
Qy 3552 CTGCCACTGTGAGCTGCGCGCAGCCTGCGCCACTTCTGACACAGCCTCCAGGGCTGCG 3611
Db 143762 CTGCCACTGTGAGCTGCGCGCAGCCTGCGCCACTTCTGACACAGCCTCCAGGGCTGCG 143703
Qy 3612 TGGGAGCATTTGGGGGCTCATGCGACCTCTGGGCTACCCACTGCGCCAGCGCTGCGTG 3671
Db 143702 TGGGAGCATTTGGGGGCTCATGCGACCTCTGGGCTACCCACTGCGCCAGCGCTGCGTG 143643
Qy 3672 GGAAGTGAACCTTGAAGGCTGAGGCTGCGGCTGCGCAGTGAATCTTCCAGAAAGATGGAAG 3731
Db 143642 GGAAGTGAACCTTGAAGGCTGAGGCTGCGGCTGCGCAGTGAATCTTCCAGAAAGATGGAAG 143583
Qy 3732 ACTTCTGGCTGCTGAAGGAGCTGAGACCTGCGCTGCGGCTGCGGCGCAAGCACTTCAAC 3791
Db 143582 ACTTCTGGCTGCTGAAGGAGCTGAGACCTGCGCTGCGGCTGCGGCGCAAGCACTTCAAC 143523
Qy 3792 GGCTCAAGAAAGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 3851
Db 143522 GGCTCAAGAAAGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 143463
Qy 3852 GCTTCTGACTTCTGACTTCT 3911
Db 143462 GCTTCTGACTTCTGACTTCT 143403
Qy 3912 AGAGCCAGCCCTGTATGCAACACCTGTGTGAGCGCAGGAGCAGAAAGCTGTGAGCCTCTGG 3971
Db 143402 AGAGCCAGCCCTGTATGCAACACCTGTGTGAGCGCAGGAGCAGAAAGCTGTGAGCCTCTGG 143343
Qy 3972 CCCTTCTGAGCCGCTGGGCGTGTGATGCGATCAGCCCTGTCTCTCTCTCTCTCTCTCTCT 4031
Db 143342 CCCTTCTGAGCCGCTGGGCGTGTGATGCGATCAGCCCTGTCTCTCTCTCTCTCTCTCTCT 143283
Qy 4032 AAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCGCTGTCTCTCTCTCTCTCTCTCTCTCT 4091
Db 143282 AAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCGCTGTCTCTCTCTCTCTCTCTCTCT 143223
Qy 4092 AGTCATGCTCAGGAGGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGT 4151
Db 143222 AGTCATGCTCAGGAGGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGT 143163
Qy 4152 GCTTCTTCCCTACCACTTGGCGAGTGGCCAGCCAGCCCTCAGGTGGGACATCTGGAGGG 4211
Db 143162 GCTTCTTCCCTACCACTTGGCGAGTGGCCAGCCAGCCCTCAGGTGGGACATCTGGAGGG 143103
Qy 4212 CAGGGGTTGAGGGGCCACACACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4271
Db 143102 CAGGGGTTGAGGGGCCACACACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 143043
Qy 4272 ACTCTCTCTGAGTGGTGTGTCT 4331
Db 143042 ACTCTCTCTGAGTGGTGTGTCT 142983

Qy 4332 AAAAACAATGTTGGCAATTTCTACAAAAAGAGATGAGATTAAACAGTGCAGGGTTGGGG 4391
Db 142982 AAAAACAATGTTGGCAATTTCTACAAAAAGAGATGAGATTAAACAGTGCAGGGTTGGGG 142923
Qy 4392 TCTGCAATGAGGTGCGCTATAAACACAGAGAAATACCTGAAAGCAGAGGGCAGGGA 4451
Db 142922 TCTGCAATGAGGTGCGCTATAAACACAGAGAAATACCTGAAAGCAGAGGGCAGGGA 142863
Qy 4452 CAGACACAGACAGACCCAGGAGTCTCCAAAGCAGAGTGGCAAAACAAACCCAGAGCTGA 4511
Db 142862 CAGACACAGACAGACCCAGGAGTCTCCAAAGCAGAGTGGCAAAACAAACCCAGAGCTGA 142803
Qy 4512 GCATCAGGACCTTCCCTCGAATTTGTCTCCAGTATTACGGTGCCTCTCTCTGCCCCCTT 4571
Db 142802 GCATCAGGACCTTCCCTCGAATTTGTCTCCAGTATTACGGTGCCTCTCTCTGCCCCCTT 142743
Qy 4572 TCCAGAGGTATCTGTGGTGTCCAGGCTGGGAGGCGCAACCATAGCCACACACAGGATT 4631
Db 142742 TCCAGAGGTATCTGTGGTGTCCAGGCTGGGAGGCGCAACCATAGCCACACACAGGATT 142683
Qy 4632 TCCTGAAAGTTTACAAATGCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCCAAGGCC 4691
Db 142682 TCCTGAAAGTTTACAAATGCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCCAAGGCC 142623
Qy 4692 TGCCCCCAGGCCCCACCACTCATGACTTAAGTGTGTGTATTAATTTATTTATTTT 4751
Db 142622 TGCCCCCAGGCCCCACCACTCATGACTTAAGTGTGTGTATTAATTTATTTATTTT 142563
Qy 4752 GAGATGTTATTTATTTAGATGATTTATTTAGCAGATTTCTATTCTTGTATTAACAATAA 4811
Db 142562 GAGATGTTATTTATTTAGATGATTTATTTAGCAGATTTCTATTCTTGTATTAACAATAA 142503
Qy 4812 AATGCTGCCCCAGAACTTAGTCTCTTTGCCAGCCTCACCCCTCTCTGCTCATCAGA 4871
Db 142502 AATGCTGCCCCAGAACTTAGTCTCTTTGCCAGCCTCACCCCTCTCTGCTCATCAGA 142443
Qy 4872 CTCTTGCCACCCCTGGTCCCACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4931
Db 142442 CTCTTGCCACCCCTGGTCCCACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 142383
Qy 4932 GGAAGAGCCCTTCTCTCCCGCACTGGGCGATGGGCGCACTCAGACTTACCCACTG 4991
Db 142382 GGAAGAGCCCTTCTCTCTCCCGCACTGGGCGATGGGCGCACTCAGACTTACCCACTG 142323
Qy 4992 CTGCTGCCACCAACCCCTTGTATCCCTCAGTCTCTCCCAACAGCTTCTGTCCACCCCA 5051
Db 142322 CTGCTGCCACCAACCCCTTGTATCCCTCAGTCTCTCCCAACAGCTTCTGTCCACCCCA 142263
Qy 5052 GGTTCCTCTCACCCGACCTTTGCTAAGTCTTCTCTCA 5087
Db 142262 GGTTCCTCTCACCCGACCTTTGCTAAGTCTTCTCTCA 142227

RESULT 6
AP02437 135116 bp DNA linear HTG 06-JUN-2000
LOCUS Homo sapiens chromosome 11 clone RP11-678D20 map 11q13, WORKING
DEFINITION DRAFT SEQUENCE, 41 unordered pieces.
ACCESSION AP02437.1 GI:8307741
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-678D20.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 135116)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 135,116 genomic DNA of 11q13
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 135116)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail: hattori@gsc.riken.go.jp,
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924)

COMMENT

----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: RP11-678D20
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 109313 bases at least Q40
 Consensus quality: 121923 bases at least Q30
 Consensus quality: 127981 bases at least Q20
 Insert size: 131116; sum-of-contigs
 Quality coverage: 4.18x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
 41 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

1 7749 contig of 7749 bp in length
 7850 15108 contig of 7259 bp in length
 15209 23728 contig of 8520 bp in length
 23829 29302 contig of 5474 bp in length
 29403 36171 contig of 6769 bp in length
 36272 42123 contig of 5852 bp in length
 42224 48172 contig of 5949 bp in length
 48273 52700 contig of 4428 bp in length
 52801 57587 contig of 4787 bp in length
 57688 61624 contig of 3937 bp in length
 61725 67756 contig of 3032 bp in length
 67857 71834 contig of 3978 bp in length
 71935 76800 contig of 4866 bp in length
 76901 81069 contig of 4169 bp in length
 81170 81932 contig of 763 bp in length
 82033 86213 contig of 4181 bp in length
 86314 89537 contig of 3224 bp in length
 89638 92372 contig of 2735 bp in length
 92473 94979 contig of 2507 bp in length
 95080 96272 contig of 1193 bp in length
 96373 98109 contig of 1737 bp in length
 98210 101030 contig of 2821 bp in length
 101131 103760 contig of 2630 bp in length
 103861 106330 contig of 2470 bp in length
 106431 108601 contig of 2171 bp in length
 108702 110440 contig of 1739 bp in length
 110541 112280 contig of 2033 bp in length
 112381 114885 contig of 1740 bp in length
 114986 117018 contig of 2029 bp in length
 117119 118839 contig of 1721 bp in length
 118940 120968 contig of 1062 bp in length
 121069 123048 contig of 1879 bp in length
 123048 124231 contig of 1396 bp in length
 124332 125727 contig of 1365 bp in length
 125828 127192 contig of 1727 bp in length
 127293 129019 contig of 1626 bp in length
 129120 130181 contig of 1202 bp in length
 130282 131483 contig of 1012 bp in length
 131584 132595 contig of 1191 bp in length
 132696 133886 contig of 1191 bp in length

133987 135116 contig of 1130 bp in length.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 7749: contig of 7749 bp in length
 7750 7849: gap of 100 bp
 7850 15108: contig of 7259 bp in length
 15109 15208: gap of 100 bp
 15209 23728: contig of 8520 bp in length
 23729 23828: gap of 100 bp
 23829 29302: contig of 5474 bp in length
 29303 29402: gap of 100 bp
 29403 36171: contig of 6769 bp in length
 36172 36271: gap of 100 bp
 36272 42123: contig of 5852 bp in length
 42124 42223: gap of 100 bp
 42224 48172: contig of 5949 bp in length
 48173 48272: gap of 100 bp
 48273 52700: contig of 4428 bp in length
 52701 52800: gap of 100 bp
 52801 57587: contig of 4787 bp in length
 57588 57687: gap of 100 bp
 57688 61624: contig of 3937 bp in length
 61625 61724: gap of 100 bp
 61725 67756: contig of 6032 bp in length
 67757 67856: gap of 100 bp
 67857 71834: contig of 3978 bp in length
 71835 71934: gap of 100 bp
 71935 76800: contig of 4866 bp in length
 76801 76900: gap of 100 bp
 76901 81069: contig of 4169 bp in length
 81070 81169: gap of 100 bp
 81170 81932: contig of 763 bp in length
 81933 82032: gap of 100 bp
 82033 86213: contig of 4181 bp in length
 86214 86313: gap of 100 bp
 86314 89537: contig of 3224 bp in length
 89538 89637: gap of 100 bp
 89638 92372: contig of 2735 bp in length
 92373 92472: gap of 100 bp
 92473 94979: contig of 2507 bp in length
 94980 95079: gap of 100 bp
 95080 96272: contig of 1193 bp in length
 96273 96372: gap of 100 bp
 96373 98109: contig of 1737 bp in length
 98110 98209: gap of 100 bp
 98210 101030: contig of 2821 bp in length
 101031 101130: gap of 100 bp
 101131 103760: contig of 2630 bp in length
 103761 103860: gap of 100 bp
 103861 106330: contig of 2470 bp in length
 106331 106430: gap of 100 bp
 106431 108601: contig of 2171 bp in length
 108602 108701: gap of 100 bp
 108702 110440: contig of 1739 bp in length
 110441 110540: gap of 100 bp
 110541 112280: contig of 1740 bp in length
 112281 112380: gap of 100 bp
 112381 114885: contig of 2029 bp in length
 114886 114985: gap of 100 bp
 114986 117018: contig of 2033 bp in length
 117019 117118: gap of 100 bp
 117119 118839: contig of 1721 bp in length
 118840 118939: gap of 100 bp
 118940 120968: contig of 2029 bp in length
 120969 121068: gap of 100 bp
 121069 122947: contig of 1879 bp in length
 122948 123047: gap of 100 bp

Sun Feb 2 08:32:00 2003

```
* 123048 124231: contig of 1184 bp in length
* 124232 124331: gap of 100 bp
* 124332 125727: contig of 1396 bp in length
* 125728 125827: gap of 100 bp
* 125828 127192: contig of 1365 bp in length
* 127193 127292: gap of 100 bp
* 127293 129019: contig of 1727 bp in length
* 129020 129119: gap of 100 bp
* 129120 130181: contig of 1062 bp in length
* 130182 130281: gap of 100 bp
* 130282 131483: contig of 1202 bp in length
* 131484 131583: gap of 100 bp
* 131584 132595: contig of 1012 bp in length
* 132596 132695: gap of 100 bp
* 132696 133886: contig of 1191 bp in length
* 133887 133986: gap of 100 bp
* 133987 135116: contig of 1130 bp in length.
```

FEATURES
Location/Qualifiers

```
1. .135116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q13"
/clone="RP11-678D20"
1. .7749
/notes="assembly_fragment"
7850. .15108
/notes="assembly_fragment"
15209. .23728
/notes="assembly_fragment"
23829. .29302
/notes="assembly_fragment clone_end:SP6 vector_side:right"
29403. .36171
/notes="assembly_fragment"
36272. .42123
/notes="assembly_fragment"
42224. .48172
```

```
Query Match 90.5%; Score 4605; DB 2; Length 135116;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 4682; Conservative 0; Mismatches 35; Indels 6; Gaps 5;

Qy 367 TATATCTACCTAGCCAAACCCCTTAAGTCTTGGCCAGTCCAAAGTGTCCACTGAATCCT 426
Db 15209 TATATCTACCTAGCCAAACCCCTTAAGTCTTGGCCAGTCCAAAGTGTCCACTGAATCCT 15268

Qy 427 CACCTTGGTCTCACTGAAATCCCAAGAAAGCATATTTCCCACTGGCCACATCCCTCC 486
Db 15269 CACCTTGGTCTCACTGAAATCCCAAGAAAGCATATTTCCCACTGGCCACATCCCTCC 15328

Qy 487 TTACAGACCCCAACCCCTGGCTCTGGACTCTGGTATCTGGATGTCCTGATGTCCTGAG 546
Db 15329 TTACAGACCCCAACCCCTGGCTCTGGACTCTGGTATCTGGATGTCCTGAGTCTGAG 15386

Qy 547 TGCATCAGCCCAACCAAGCCGACTCGTCAATGCAATGCACTCTCTCCCTTCCTGTCCTCC 606
Db 15387 TGCATCAGCCCAACCAAGCCGACTCGTCAATGCAATGCACTCTCTCCCTTCCTGTCCTCC 15446

Qy 607 TTGAGGCTGATGGAAGGCTCATTTGAAGTCCMACTTTTCCCACTTAACCAACCAAGAAC 666
Db 15447 TTGAGGCTGATGGAAGGCTCATTTGAAGTCCMACTTTTCCCACTTAACCAACCAAGAAC 15506

Qy 667 GGGGTGAACTCCACACTGCGACCTGCTTCCCTGAGAGTGAGCACTAAATCTCTTCAATCT 726
Db 15507 GGGGTGAACTCCACACTGCGACCTGCTTCCCTGAGAGTGAGCACTAAATCTCTTCAATCT 15566

Qy 727 AACCCCAACCCCTACACTTTCCCACTCAAGGAATCATCTCTAGATATATACCAAACTAAG 786
Db 15567 AACCCCAACCCCTACACTTTCCCACTCAAGGAATCATCTCTAGATATATACCAAACTAAG 15626

Qy 787 CCCCATAGGACGCGCCGACCTAGTGTCTTAACCTTATACCTTGTCTTCTTATGGTGAGT 846
Db 15627 CCCCATAGGACGCGCCGACCTAGTGTCTTAACCTTATACCTTGTCTTCTTATGGTGAGT 15686
```

```
Qy 847 CTGTTCTTGGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 905
Db 15687 CTGTTCTTGGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 15746

Qy 906 GCCAGCTCTGACATGTGTCTCTCCACCCCTCTGACCTCCCTCAAGCTGCACTGGGACTG 965
Db 15747 GCCATCTCTGACATGTGTCTCTCCACCCCTCTGACCTCCCTCAAGCTGCACTGGGACTG 15806

Qy 966 GAAAGCTGGCAGGAAGCTAGGGTACAACTGGGAACACAGCAGCTGAGCTGCCCTA 1025
Db 15807 GAAAGCTGGCAGGAAGCTAGGGTACAACTGGGAACACAGCAGCTGAGCTGCCCTA 15866

Qy 1026 GGCCTGGCCCGCTCTCTCATGTACACATATACATATGTTGGCACACACAGTGGGACA 1085
Db 15867 GGCCTGGCCCGCTCTCTCATGTACACATATACATATGTTGGCACACACAGTGGGACA 15926

Qy 1086 CATGCAAAAGACTCTCTCAGCTGACACACAGATCCATCTCAAGTATCTACTGATAGACA 1145
Db 15927 CATGCAAAAGACTCTCTCAGCTGACACACAGATCCATCTCAAGTATCTACTGATAGACA 15986

Qy 1146 CTCATGCGTGGCAAGTCTCTCATCTCAACATACACATGCTCTCTCTCTCTCTCTCTCTCT 1205
Db 15987 CTCATGCGTGGCAAGTCTCTCATCTCAACATACACATGCTCTCTCTCTCTCTCTCTCTCT 16046

Qy 1206 GCCAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1265
Db 16047 GCCAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16106

Qy 1266 CCCCACCCAGCCAGGTGGGACACACACCTGAGGGGCTGCCAGCTGCTTCCCGGTGTG 1325
Db 16107 CCCCACCCAGCCAGGTGGGACACACACCTGAGGGGCTGCCAGCTGCTTCCCGGTGTG 16166

Qy 1326 GGCCTGGCCCGCTCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1385
Db 16167 GGCCTGGCCCGCTCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16226

Qy 1386 TAGCTGCTCTGTCAGCTGCTCTGGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
Db 16227 TAGCTGCTCTGTCAGCTGCTCTGGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16286

Qy 1446 GGCACCCAGGCTCTGGCCCTCTCATCCAGAAACCTATGACCTCACCCGCTACCTGGAGC 1505
Db 16287 GGCACCCAGGCTCTGGCCCTCTCATCCAGAAACCTATGACCTCACCCGCTACCTGGAGC 16346

Qy 1506 ACCAACTCCGAGCTTGGCTGGGACCTATGTAGTATCCAGGTAGGAATCTGGGAGTTG 1565
Db 16347 ACCAACTCCGAGCTTGGCTGGGACCTATGTAGTATCCAGGTAGGAATCTGGGAGTTG 16406

Qy 1566 GGGAGGAGTGGAGGTGGGGAAAGACAGTCTTAACCTGGAGGCTTCTGGTAAATGATG 1625
Db 16407 GGGAGGAGTGGAGGTGGGGAAAGACAGTCTTAACCTGGAGGCTTCTGGTAAATGATG 16466

Qy 1626 GGGTGGAGGAGGCTCTTTTGGCTCTCCACAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 1685
Db 16467 GGGTGGAGGAGGCTCTTTTGGCTCTCCACAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 16526

Qy 1686 CCCCCTTAGTGGCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1745
Db 16527 CCCCCTTAGTGGCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16586

Qy 1746 GCCTCGACCCCGCTTGGCCCAATTCGCCCACTTGGCTGCGAGCCCGCCCTCTCTCTCTCT 1805
Db 16587 GCCTCGACCCCGCTTGGCCCAATTCGCCCACTTGGCTGCGAGCCCGCCCTCTCTCTCTCT 16646

Qy 1806 CTGGGGGCGGGGAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1865
Db 16647 CTGGGGGCGGGGAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16706

Qy 1866 GGTGGGTGGGACAGAGGGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1925
Db 16707 GGTGGGTGGGACAGAGGGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16766
```


QY 1926 CCAGACTGGGGCCCTGCTGCTCTGGAACCCAGGGGCTCCCTCCGCTGCTCCCTCCCAT 1985
Db 16767 CCAGACTGGGGCCCTGCTGCTCTGGAACCCAGGGGCTCCCTCCGCTGCTCCCTCCCAT 16826
QY 1986 CTTAGCTGGGCCCTCCTAGGGGGCTCATGGGGGAGGGAGCTGAGGGAACCCAGGCACTA 2045
Db 16827 CTTAGCTGGGCCCTCCTAGGGGGCTCATGGGGGAGGGAGCTGAGGGAACCCAGGCACTA 16886
QY 2046 GTGGCAGGGGTTTAGGGTGTGATGGAGGTTATGCTGTAAAGATTGGGGGTGGTCCAG 2105
Db 16887 GTGGCAGGGGTTTAGGGTGTGATGGAGGTTATGCTGTAAAGATTGGGGGTGGTCCAG 16946
QY 2106 AGGTGTTTCAGAGACCCAGGAGAGAGGAGGAGGTTGAGAGCCGAGGCAACATGGG 2165
Db 16947 AGGTGTTTCAGAGACCCAGGAGAGAGGAGGAGGTTGAGAGCCGAGGCAACATGGG 17005
QY 2166 GAAACGGGCCCTCTTCCGCTGCTCTCTTCCATCCAGACCTTACTCTGGAGCCAGG 2225
Db 17006 GAAACGGGCCCTCTTCCGCTGCTCTCTTCCATCCAGACCTTACTCTGGAGCCAGG 17065
QY 2226 GAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2285
Db 17066 GAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 17125
QY 2286 ATTAGTTTGTCTCTGTGCTTGTGAGGTTTGTGAGGTTTGTGAGGTTTGTGAGGTTT 2345
Db 17126 ATTAGTTTGTCTCTGTGCTTGTGAGGTTTGTGAGGTTTGTGAGGTTTGTGAGGTTT 17185
QY 2346 GGCCTCTCCCTTGTCT 2405
Db 17186 GGCCTCTCCCTTGTCT 17245
QY 2406 GGG-CCCCAGCGCTCCCGGAGGTTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 2464
Db 17246 GGGCCCCAGCGCTCCCGGAGGTTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 17305
QY 2465 TGTCTTTCATAGCTTCT 2524
Db 17306 TGTCTTTCATAGCTTCT 17365
QY 2525 AACCTGTCTCTATCAGCTGAACCTTTCTTGGAGTGTAGTGAATCCGCTCTCTCC 2584
Db 17366 AACCTGTCTCTATCAGCTGAACCTTTCTTGGAGTGTAGTGAATCCGCTCTCTCC 17425
QY 2585 CAGCCCTCAGCTGGGGCTGGTGTGAGGGGCAAGGAGGAGGAGGAGGAGGAGGAGG 2644
Db 17426 CAGCCCTCAGCTGGGGCTGGTGTGAGGGGCAAGGAGGAGGAGGAGGAGGAGGAGG 17485
QY 2645 GCGACT 2704
Db 17486 GCGACT 17545
QY 2705 CTCTAGTTCCGACCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2764
Db 17546 CTCTAGTTCCGACCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17605
QY 2765 TCTACACCTCTGCT 2824
Db 17606 TCTACACCTCTGCT 17665
QY 2825 GCTTCCAGGCCCAGCAATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2884
Db 17666 GCTTCCAGGCCCAGCAATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17725
QY 2885 ACAGTGTGTTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2944
Db 17726 ACAGTGTGTTTGAAGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17785
QY 2945 GCTGCTCTCTGAACT 3004
Db 17786 GCTGCTCTCTGAACT 17845
QY 3005 TCCAGTCTCTCAGGCT 3064

Db 17846 TCCAGTCTCTCAGGCT 17905
QY 3065 GGTTCGCCACCCAGCAATCAGCAGCTCTCTGAGCGCTCTGTCGCTCTCTCTCTCTCT 3124
Db 17906 GG-TTGGCCACCCAGCAATCAGCAGCTCTCTGAGCGCTCTGTCGCTCTCTCTCTCT 17964
QY 3125 TCCCTTTTCTACGCT 3184
Db 17965 TCCCTTTTCTACGCT 18024
QY 3185 GCATACAAATGATATCTCTTATTGAAAAAATCAGGGAGGCAATGAAAAAAGAGGCTAGC 3244
Db 18025 GCATACAAATGATATCTCTTATTGAAAAAATCAGGGAGGCAATGAAAAAAGAGGCTAGC 18084
QY 3245 ATGGAGACAGGGCAGTGTCAAGGGGACACAAAAAATAGAAAACTTTGGGAGCAGGTATCTC 3304
Db 18085 ATGGAGACAGGGCAGTGTCAAGGGGACACAAAAAATAGAAAACTTTGGGAGCAGGTATCTC 18144
QY 3305 CTTGTGTGTGAGCCAGCGCTCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3364
Db 18145 CTTGTGTGTGAGCCAGCGCTCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 18204
QY 3365 CTGAACCTACTCTGGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGGCA 3424
Db 18205 CTGAACCTACTCTGGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGGGCA 18264
QY 3425 GAGACTCTGCCAGCGGCACTGTTCATCTGGAGGTGTGGCGAAGCTCAATGACAAACTG 3484
Db 18265 GAGACTCTGCCAGCGGCACTGTTCATCTGGAGGTGTGGCGAAGCTCAATGACAAACTG 18324
QY 3485 GGGCTGACCCAGAACTACGAGCCCTACAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 3544
Db 18325 GGGCTGACCCAGAACTACGAGCCCTACAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 18384
QY 3545 CGTCAGGCTGCACTGTCTGAGCTGCGCGCAGCTCTGGGCCCTCTCTCTCTCTCTCTCT 3604
Db 18385 CGTCAGGCTGCACTGTCTGAGCTGCGCGCAGCTCTGGGCCCTCTCTCTCTCTCTCTCT 18444
QY 3605 GGCCTGCTGGGAGCACTTGGGGGCTCAAGGAGCTCTGGGCTACCCCTCTCTCTCTCTCT 3664
Db 18445 GGCCTGCTGGGAGCACTTGGGGGCTCAAGGAGCTCTGGGCTACCCCTCTCTCTCTCTCT 18504
QY 3665 CTGCTCTGGGACTGAACCCACTTTGAGCTCTCTGGCCCTGCGCCAGAGTCTCTCTCTCT 3724
Db 18505 CTGCTCTGGGACTGAACCCACTTTGAGCTCTCTGGCCCTGCGCCAGAGTCTCTCTCTCT 18564
QY 3725 ATGAGCACTTCTGCTGTGAAGGAGCTGAGGAGCTCTGGGCTCTGGGCTCTGGGCTCT 3784
Db 18565 ATGAGCACTTCTGCTGTGAAGGAGCTGAGGAGCTCTGGGCTCTGGGCTCTGGGCTCT 18624
QY 3785 TTCAACCGGCTCAAGAAAGAGTGCAGCCCTCCAGCAGCTGAGTCACTCTGAGCTGGG 3844
Db 18625 TTCAACCGGCTCAAGAAAGAGTGCAGCCCTCCAGCAGCTGAGTCACTCTGAGCTGGG 18684
QY 3845 GCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3904
Db 18685 GCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 18744
QY 3905 CTTTGTGAGACGAGCCCTGTATGCCAACCTTGTGAGCCAGGAGACAGAACTGTGAG 3964
Db 18745 CTTTGTGAGACGAGCCCTGTATGCCAACCTTGTGAGCCAGGAGACAGAACTGTGAG 18804
QY 3965 CTTCTGGCCCTTCTCTGAGCCGCTGGGCTGTGAGTGCAGTCACTCTCTCTCTCTCTCT 4024
Db 18805 CTTCTGGCCCTTCTCTGAGCCGCTGGGCTGTGAGTGCAGTCACTCTCTCTCTCTCTCT 18864
QY 4025 CTTCTCCAAAGGCTCTACCGAGCTGGGAGGAGTACAGTAGGCCCTGTCTCTCTCTCT 4084
Db 18865 CTTCTCCAAAGGCTCTACCGAGCTGGGAGGAGTACAGTAGGCCCTGTCTCTCTCTCTCT 18924
QY 4085 TACAGGAAGTCACTGCTCGAGGAGTGTGAAGTGTGAGTGTGAGGAGGCTCTCTCTCT 4144

Assembly program: Phrap; version 0.960731
Consensus quality: 257668 bases at least
Consensus quality: 268459 bases at least
Consensus quality: 268627 bases at least
Insert size: 242000; agarose-fp
Insert size: 268755; sum-of-contigs
Quality coverage: 10.4 in Q20 bases; agarose
Quality coverage: 9.4 in Q20 bases; sum-of-

QY	1076	CAGTGGCACACATGCCAAGACTCTCTCAGCTGCACACAGATCCATTCTCAAGTATCTTA	1135
Db	149146	-----CATATTTTGGCAGATACATGCCAAGACTTTACGTAAATAGCAGTCAAGTATCTTA	149199
QY	11136	CTGATAGACACTCATGCGTGCCAAGTCTCATCTCTCAACATACACATGCCTCTCTTTCT	11195
Db	149200	CTGATAGACAAATGCGAGATTGTCT-----TTGCATACATACCTCCACAT	149248
QY	11196	CTCCGCTTTGCCAGGAGTGTTCCTCCCTCTCTCCATCCCTCTGCTCCCATCTGTGTGCC	1255
Db	149249	ACGCTCTTGACACAGAAAGTCTTTCCTCTCTCCACCCCTCCATTTCCTCACTGGTCTCT	149308
QY	1256	CACCCTCACCCCCC-ACCACGCCAAGGTGGGACAGACACTGAGGGGCTGCCAGTGC	1314
Db	149309	CTCCCTCACCCCCCTTCTTAGCCCAAGGTGGGACAGACACTGAGGGGCTGCCAGTGT	149368
QY	1315	TTCCCCGTGTGGCCCGGG-CGCGCTCTATGCTTCTCGTCCATCTCTGCCACAGGGACT	1373
Db	149369	TTCCCTATTTTGGSCCTGGGCCCAACCTCATACTTCTTGCCCATCTATTACAGGGGACT	149428
QY	1374	CGTGGGGAGTGTTAGCGTGCCTGTGACGCTGTCTGGCACTCCCTGCAGTGCAGCTC	1433
Db	149429	CGTGGGGAGTGTTAGCTTGCTTATGACACGCTGTGTGGCACTCCCTGCAGTGCAGCTC	149488
QY	1434	TCAATCGCACAGGGACCCAGGSCCTGGCCCTCTCATTCAGAAACCTATGACCTCAACC	1493
Db	149489	TTAATCGCACAGGAGATCCAGGCCCTGGCCCTCCATCCAGAAACCTATGACCTCAACC	149548
QY	1494	GCTACTGAGACACAACTCCGAGCTTGCTGGGACCTATGTGAGTATCCAGCGTAGGA	1553
Db	149549	GCTACTGAGCATCAACTCCGAGCTTAGCTGGGACCTACGTGAGTATCCCTTTGSCA	149608
QY	1554	ATCTGGGAGTTGGGGAGGAGTGTGGGAAAGACAGCTCCTAACCTGGAGGGTTC	1613
Db	149609	A-----AGGGAGGCMAGGCTTTGGGGAAGAGAGCATGGGGCAGGGGTCC	149656
QY	1614	TGGTAAATGATGGGTGAGGAGGGCTCTTTGGCTCCCAACAGTCCCTCTGTGTGCTA	1673
Db	149657	TAGTGAATGATGGGTGATGAGAGGCCCTTTGGTCCACATAGCTTCTCCATCTGGCCTA	149716
QY	1674	TCCTCTGCCCTTCCCTCTTAGTGTGGCCCCCACCCTTCCCACCTCTGGCCCCCAGGACTAG	1733
Db	149717	TCCTCTGCCCTTCCCTCTCAGTGTGCCCC-----CCCCAATCTTGGCCCCCAGGAGTAG	149769
QY	1734	GCATGTGGCAGGCTTCGCACCCGCTTGGCCCATTTGCCCCACTGGCTGCAGCCACGACC	1793
Db	149770	GCATGTGGCAGGCTTCGAACCCGCTTGGCCCATTTGCCCATTTGCTGCCAGCCAGCT	149829
QY	1794	GCCGCTCCCTCCCTGGGGCCGGGAGTCTCTCTGTTTACCCGTGTGTGGTGTCTC	1853
Db	149830	GCTTGCCTCCCTCCCTGGGGCTGGGGAAGTCTCTCTGTTTACACCGTGTGTGGTGTCTC	149889
QY	1854	TTGCGCGCGCGGGTGTGGGTGGGACAGAGGGGCCCACTCCCATCGCTGCTTCCAGC	1913
Db	149890	TTGCAAGGCTGGGCTGTGTGGGATGAGAGGGCCCCACCTTCTATGCTGTGTTCAGC	149949
QY	1914	TCGCTCTGCCCCCAGACTCTGGGCCCCCTGTGCTCTGAGCCAGGGGCTCTCCCTCCGTC	1973
Db	149950	TCGC--CTACCCCCAGACTCTGGGCCCCCTGTGTTCTTGACCCATGGCCCTCCC-CTGTCTC	150006
QY	1974	TGCCTCTCCCATCTAGCTGGGCTCTTAGGGGGTCTATGGGGGAAGGGACTGTAGGGA	2033
Db	150007	TGCCTCTCCCATCTTACAGACCTCTTAGGGG-----TTGGAGGAGGGATCTACAAGAA	150063
QY	2034	ACCAGGAGTGTGGCAGGGGGTTTAGGTGTGGATGGAGGTATGTGTGAAGATTGTG	2093
Db	150064	TCTAGGCGAGTCTATGGCAAGGGGTTACGGGTATGGAGGTGTGTGTGCCATAAAGATTG	150123
QY	2094	GGGTGTGTCCAGAGGTGTTACAGAGGCCCGAGAGAGGAAGAGGGTGTGGAGAGCCG	2153
Db	150124	AGGTAGTGTTAGATGTTTACAGAGCCCGAGAGCCGAGAGAGGATCGGAGGCATGTG	150183

QY	2154	AGGACCATGGGNAACGGGCCCTCTTCCCGTGTTCTCTTCACATCCAGACCCCTAC	2213
Db	150184	AGG-ACCATTGGGAACAGG-CCTCTTCTCCGTTCTCTTCCANATCCAGACCCCTAC	150241
QY	2214	TCTGGAGCAGGGAAGAAAGGGAAGAAGGTGGCGGGGAGCTTGGCTCCAGCCCCAGGA	2273
Db	150242	TCCTGAGCCAGG-AAGAAAAGGAGAGGTGGCGGGGAGCTGGCTCCAGCCCCAGGA	150300
QY	2274	TACACGAGGAATAAGTTGTCTCTGTGTGTGCAGCTGTGACACTCCCTCGGGCCC	2333
Db	150301	TACACGAGGAATAAGTTGTGTCTCTGTGTGTGCAGCTGTGAACTCCCTCGGGCCA	150360
QY	2334	TTGCTATCCAGAGCCCTCCCTCTGTCTCTCCCTCTTTCCAGTTATATCATCTCCCTC	2393
Db	150361	GTGCTATCCAGAGCCCTCGCCCTCTCTCT--TTCCCTCCAGTTGCAATCTCCAC	150417
QY	2394	ATCCCTTTCCCTGGGGCCCCA-GCGCCTCCCCGAGGGTGTGAAGGGCTCTGCCCTCTTC	2452
Db	150418	ATCCCTTTCCCTGGGGCTCAGGCGCTTCCCTGAGACAATTGGCAAGACTTTGCGCTCTT	150477
QY	2453	CCTATACATGCTGTCTCCATAGACCTTCCTCTCTCTCATCATGAGACTGCCCTCCATT	2512
Db	150478	CCGATGTCTCTCTGCTCCCATGCCCCCTCCCTCTCTCT-CTCTTAGATTCTCTCAGTT	150536
QY	2513	CTCTCTCTTCCAAACCCCTGCTCTATCAGCTGAACCTCTTTCCGAGTGTAGTAGTA	2572
Db	150537	CTCTCTCTT-TACAGCTCTGCTCTTAGGAACTGAGTTCTCTTTTGGG-----GTAAAGTA	150590
QY	2573	CCCGTCTCTCCAGCCCTCAGCTGTGTGGGCTGGGTGTGTGACGGCAATGGGGCTC	2632
Db	150591	CATTGCTATCTTGTCTCTTGGCTGTGTGAGGAGAGTTTGTGTGTGACAGGCTCCCT	150650
QY	2633	TGGTTCCAAATGGGCCACTCTCATCTCTCTTGTCTCTTGTGTGCAAAAAACCTTTGCTTCA	2692
Db	150651	TTGCCCTTTGATATGACTCTCTCTGCGTGTGTTC--TGCAAAAAACCTCTACTTGA	150709
QY	2693	CTCCACTGCCCTCTCTAGTTCCGACCCCTTTTCTCTCTCTGCTTTCCCTGCAAAATTTC	2752
Db	150710	GTCGTCTAACCTCTCTAACTC---GACCTTTCTCTCTAACTTCTCTGCAAAATTTC	150765
QY	2753	TCCAAAGGAGTGTCTACACCTCTGCTCCACTTCTCTCCAGCCACTCACTTCTTAACC	2812
Db	150766	TGAAAGGAGTCTCTACACCTCTGCTCCACTTCTCTCCACCCACTCACCTTGTAAACC	150825
QY	2813	CCCTGCANCTGGCTTCCAGGCCCCAGCAATGGTTCTCTCCAAAGTCTGTGAGCCACTCC	2872
Db	150826	CCCTGCAATCTGGCTTCCAGGCCACGGCAAT-CACTCTCCAGGTCATCAGGCATCTCC	150884
QY	2873	TTTGCCAAAGCCGACAGTGTTTGAAGGCTCAFTCTCTTGTGTCTGTGTGAGGCCACA	2932
Db	150885	TTTCTCA----CCAAACCCGACAGCAFTTTCTGGGCTCACTCTCATGTCTGCAACTGCC	150940
QY	2933	CTGCTGAGGCTGTCTGCTTCTGAACTCTCTCTCTCTTGTGTCTCTGCACTCTCTGGGC	2992
Db	150941	CTGCGAGTCTGTCTGCGCTCTGAGCTCTTC-----CTGGGCTGCATCTCAGAGGCC	150995
QY	2993	ACCTTCTACTCTCCAGCTCTCTCAGGCTCTCTCTCTCTCTCTCTGCTGCCCCCAAGCGG	3052
Db	150996	ACCTTCTACTCTTCCAGCTGTGTCTCTGCTCTCTC-GGCTCTCTCTCTCTCTGCTCT	151054
QY	3053	GCACTCTCCAAAGTTTGCCCAACCCAGCCAAATCAGCACTGTCTCTCTGAGGCTTGTGTGC	3112
Db	151055	GCCCTGTGTGGGCAATTCTCAACCCAGACAGTCAGACAGTCCCAACCCGAGTGTCTCTTC	151114
QY	3113	GTCTCTCTCTCTCTTTTCTAGGCTCTCCATTTGGAGAGCTCACACCGCCACTGTCTT	3172
Db	151115	GTGGTTTCTCTC-----TCTCTGCATCTCTCCATAGAGAACTCATCACC-CCACTGTCTT	151167
QY	3173	CAACTGTCACTGCATAAAATGATATCTTTATTTGAAAAAAGCTCAGGGAGGCCATGAACA	3232
Db	151168	CACCTGTCTCTCCATGCAATGATACACTTTGCTGGAAAACT-----151210	
QY	3233	AAGAGCCTTAGCATGTGGAGACAGGGCCAGTGTGAGGGGACAAAAATAGAAACTTTGGG	3292

Db 767 CACTTTGTGAGAGCCGCTGTATGCGCAACACCTGTGTGAGCCAGAGAGACAGAAAGCTGTG 826
QY 3963 AGCCTCTGGCCCTTCTCTGACCGGCTGGCGCTGTGATCGATCAGCCCTGTCTCTCC 4022
Db 827 AGCCTCTGGCCCTTCTCTGACCGGCTGGCGCTGTGATCGATCAGCCCTGTCTCTCC 886
QY 4023 CACCTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTT 4082
Db 887 CACCTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTT 946
QY 4083 TCTACAGGAAGTCACTGCTCAGGAGGTGTGAGTGGTTCAGGTGGTGGAGGCGCTCA 4142
Db 947 TCTACAGGAAGTCACTGCTCAGGAGGTGTGAGTGGTTCAGGTGGTGGAGGCGCTCA 1006
QY 4143 TGGCT 4202
Db 1007 TGGCT 1066
QY 4203 TCTGGAGGCGAGGCTGTGAGGCGCCACACACACATGCTCTCTCTCTCTCTCTCTCT 4262
Db 1067 TCTGGAGGCGAGGCTGTGAGGCGCCACACACACATGCTCTCTCTCTCTCTCTCTCT 1126
QY 4263 GGTCTGCCCATCT 4322
Db 1127 GGTCTGCCCATCT 1186
QY 4323 CAATTACAGAAACAAACATGTGGTGGCAATCTTACACAAAAGAGATGAGATTAAACAGTCA 4382
Db 1187 CAATTACAGAAACAAACATGTGGTGGCAATCTTACACAAAAGAGATGAGATTAAACAGTCA 1246
QY 4383 GGTGTGGGCTGTGCAATGTGGGCTGCTTATTAACACAGAGAGAAAATCTGAAAGCACAG 4442
Db 1247 GGTGTGGGCTGTGCAATGTGGGCTGCTTATTAACACAGAGAGAAAATCTGAAAGCACAG 1306
QY 4443 GGGCAGGACAGACACAGACCCAGAGCTCTCAAGACACAGAGTGGCAACAAAC 4502
Db 1307 GGGCAGGACAGACACAGACCCAGAGCTCTCAAGACACAGAGTGGCAACAAAC 1366
QY 4503 CCGAGCTAGGATCAGGACCTTGGCTCGAATGTCTTCCAGTATTACGGTGGCTCTCTCT 4562
Db 1367 CCGAGCTAGGATCAGGACCTTGGCTCGAATGTCTTCCAGTATTACGGTGGCTCTCTCT 1426
QY 4563 TGGCCCTCTTCCAGGCTATCTGTGGTTCGAGCTGGGAGGCGCAACATAGCCACAC 4622
Db 1427 TGGCCCTCTTCCAGGCTATCTGTGGTTCGAGCTGGGAGGCGCAACATAGCCACAC 1486
QY 4623 CACAGGATTTCTGAAAGTTTACAAATGACGTAGCAATTTTGGGTGTAGGGTGGAGCTCC 4682
Db 1487 CACAGGATTTCTGAAAGTTTACAAATGACGTAGCAATTTTGGGTGTAGGGTGGAGCTCC 1546
QY 4683 CCAAGGCTCTGCCCCCAGCCCACTCATGACTTAAGTGTGTGTTAATAATTT 4742
Db 1547 CCAAGGCTCTGCCCCCAGCCCACTCATGACTTAAGTGTGTGTTAATAATTT 1606
QY 4743 ATTTATTGGAGATGTTATTATTATTAGATGATATTATTATTCAGATTTCTTCTGATT 4802
Db 1607 ATTTATTGGAGATGTTATTATTATTAGATGATATTATTATTCAGATTTCTTCTGATT 1666
QY 4803 AACAAATAAATGTTGCCCCAG 4825
Db 1667 AACAAATAAATGTTGCCCCAG 1689

RESULT 10
AX205024
LOCUS AX205024
DEFINITION Sequence 1 from Patent WO0155172.
ACCESSION AX205024
VERSION AX205024.1 GI:15394259
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 881)
Elson,G., Gauchat,J.F., Plun-Favreau,H., Chevalier,S. and Gascan,H.
Isolated complex comprising a nnt-1 protein and in addition at
least a clf-1 protein and/or a scntfr_g(a) protein
Patent: WO 0155172-A 1 02-AUG-2001;
PIERRE FABRE MEDICAMENT (FR); INSTITUT NATIONAL DE LA SANTE ET DE
LA RECHERCHE MEDICALE (INSERM) (FR)
Location/Qualifiers
1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"
174..851
/note="NNT-1"
/codon_start=1
/protein_id="CAC60175.1"
/tr_xref="GI:15394260"
/translation="MDLRAGDSWGLACLCTVLMHLPALNRTGDPGPGSIQKTY
DLTVLEHOLRLSLAGTYLNLGPPNEPDPNPLRGAETLPRATYDLEWVRSNLKLR
LTQNYEAYSHLLCYLRGNROAATAELRSLAHFCTSLQGLLGSAGVAAAGYPLPQ
PLPSTPTWTFGPAHSDFLQKMDDFLLKLEQLTWLWRSKDFNRLKKMQPPAAVTL
HLGAHGF"
BASE COUNT 158 a 318 c 246 g 159 t
ORIGIN
Query Match 10.3%; Score 525.4; DB 6; Length 881;
Best Local Similarity 99.8%; Pred. No. 2.5e-117;
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCTGGGGG 3422
Db 355 ATCTGAACCTACCTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCTGGGGG 414
QY 3423 CAGAGACTCTGCCAGGCGCACTGTGTGACTTGTGAGGTGTGGAGGTGTGGAGGCTCAATGACAAAC 3482
Db 415 CAGAGACTCTGCCAGGCGCACTGTGTGACTTGTGAGGTGTGGAGGCTCAATGACAAAC 474
QY 3483 TGGCGCTGACCCAGAACCTACAGGCGCTTACAGGCGCTTCTGTGTACTTGTGTGCGCTCA 3542
Db 475 TGGCGCTGACCCAGAACCTACAGGCGCTTCTGTGTACTTGTGTGCGCTCA 534
QY 3543 ACCGTGAGGCTGCCACTGTCTGAGCTTGGCGCGCAGCTTGGCGCGCAGCTTCTGACACAGCTCC 3602
Db 535 ACCGTGAGGCTGCCACTGTCTGAGCTTGGCGCGCAGCTTGGCGCGCAGCTTCTGACACAGCTCC 594
QY 3603 AGGCGCTCTGGGCGAGCAATTCGGGCGGTCTATGGGCGTCTGGGCTACCCACTGCCCCAGC 3662
Db 595 AGGCGCTCTGGGCGAGCAATTCGGGCGGTCTATGGGCGTCTGGGCTACCCACTGCCCCAGC 654
QY 3663 CGCTGCTGGGAGCTGAACCCCACTTGGAGCTTCTGGCGCGCTTGGCGCGCAGCTTCTCTCCAGA 3722
Db 655 CGCTGCTGGGAGCTGAACCCCACTTGGAGCTTCTGGCGCGCTTGGCGCGCAGCTTCTCTCCAGA 714
QY 3723 AGATGAGACGACTTCTGGCTGTCTGAAGGAGCTGACAGCTGGCTGTGGCGCTGGCGCAAGG 3782
Db 715 AGATGAGACGACTTCTGGCTGTCTGAAGGAGCTGACAGCTGGCTGTGGCGCTGGCGCAAGG 774
QY 3783 ACTTCACCGGCTCAAGAAGAGTGCAGGCTTCAGAGCTGCAGTGCACCTGCACCTCGG 3842
Db 775 ACTTCACCGGCTCAAGAAGAGTGCAGGCTTCAGAGCTGCAGTGCACCTGCACCTCGG 834
QY 3843 GGGCTCATGCTTCTGACTTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3889
Db 835 GGGCTCATGCTTCTGACTTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 881
RESULT 11
AX205042
LOCUS AX205042
DEFINITION Sequence 1 from Patent WO0155219.
ACCESSION AX205042
VERSION AX205042.1 GI:15394277
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 881)
TITLE Elson, G. and Gauchat, J.F.
JOURNAL Scentfr/nnt-1 fusion protein
PATENT: WO 0155219-A 1 02-AUG-2001;
PIERRE FABRE MEDICAMENT (FR)
FEATURES Location/Qualifiers
source 1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 174..851
/note="NNT-1"
/codon_start=1
/protein_id="CAC60178.1"
/db_xref="GI:15394278"
/translation="NDLRAGDSWGLACLCTVHLPAVPALNRTGDPGPSPSIOKTY
DLTRYELHQLRLAGLYNLGPPNEPFPRLGAETLPATVDLEVRSLNDKLR
LTQNYEAYSHLLCYLRGLNRQAATAEELRSLAFCFSLQGLLSIAGVMAALGYPLPQ
PLPGTEPTWTPGPAHSDFLQKMDDFWLKELQTLWRSKDFNRLKKKMQPPAAAVIL
HLGARGF"
BASE COUNT 158 a 318 c 246 g 159 t
ORIGIN
Query Match 10.3%; Score 525.4; DB 6; Length 881;
Best Local Similarity 99.8%; Pred. No. 2.5e-117;
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGGG 3422
DB 355 ATCTGAACCTACCTGGGCCCCCTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 414
QY 3423 CAGAGACTCTGCCCCAGGCCCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482
DB 415 CAGAGACTCTGCCCCAGGCCCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 474
QY 3483 TGGCGCTGAGCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGCGCTCA 3542
DB 475 TGGCGCTGAGCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGCGCTCA 534
QY 3543 ACCGTGAGGCTGACCTGCTGAGCTGGCGCGCAGCCTGGCCACTTCTGACCCAGCCTCC 3602
DB 535 ACCGTGAGGCTGACCTGCTGAGCTGGCGCGCAGCCTGGCCACTTCTGACCCAGCCTCC 594
QY 3603 AGGGCCTGCTGGGAGCAGATTGCGGGCGTATGGAGCTCTGGGCTACCACTGCCCCAGC 3662
DB 595 AGGGCCTGCTGGGAGCAGATTGCGGGCGTATGGAGCTCTGGGCTACCACTGCCCCAGC 654
QY 3663 CGTGCTGGGAGCTGAGCCCACTTGGACTCTGGCCCTGCGCCACAGTGCATCTCTCCCA 3722
DB 655 CGTGCTGGGAGCTGAGCCCACTTGGACTCTGGCCCTGCGCCACAGTGCATCTCTCCCA 714
QY 3723 AGATGACGACTTCTGGCTGTGAGGAGCTGCAGACCTGGCTGTGGCGCTGGCCCAAG 3782
DB 715 AGATGACGACTTCTGGCTGTGAGGAGCTGCAGACCTGGCTGTGGCGCTGGCCCAAG 774
QY 3783 ACTTCAACCGGCTCAAGAAGATGAGCCCTCAGAGCTGACCTGACCTGACCTGACCTGG 3842
DB 775 ACTTCAACCGGCTCAAGAAGATGAGCCCTCAGAGCTGACCTGACCTGACCTGACCTGG 834
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTTCCTTCCCTCCCTT 3889
DB 835 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTTCCTTCCCTCCCTT 881
RESULT 12
AR002595
LOCUS AR002595
DEFINITION Sequence 1 from patent US 5741772.
ACCESSION AR002595
VERSION AR002595.1 GI:3964149

KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 797)
AUTHORS Chang, M.-s.
TITLE Neurotrophic factor NNT-1
JOURNAL Patent: US 5741772-A 1 21-APR-1998;
FEATURES Location/Qualifiers
source 1..797
/organism="unknown"
BASE COUNT 139 a 297 c 218 g 143 t
ORIGIN
Query Match 10.3%; Score 523.4; DB 6; Length 797;
Best Local Similarity 99.8%; Pred. No. 7.7e-117;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGGG 3422
DB 271 ATCTGAACCTACCTGGGCCCCCTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGGG 330
QY 3423 CAGAGACTCTGCCCCAGGCCCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 3482
DB 331 CAGAGACTCTGCCCCAGGCCCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAAC 390
QY 3483 TGGCGCTGAGCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGCGCTCA 3542
DB 391 TGGCGCTGAGCCAGAACTACGAGCCCTACAGCCACTTCTGTGTTACTTGGTGCGCTCA 450
QY 3543 ACCGTGAGGCTGACCTGCTGAGCTGGCGCGCAGCCTGGCCACTTCTGACCCAGCCTCC 3602
DB 451 ACCGTGAGGCTGACCTGCTGAGCTGGCGCGCAGCCTGGCCACTTCTGACCCAGCCTCC 510
QY 3603 AGGGCCTGCTGGGAGCAGATTGCGGGCGTATGGAGCTCTGGGCTACCACTGCCCCAGC 3662
DB 511 AGGGCCTGCTGGGAGCAGATTGCGGGCGTATGGAGCTCTGGGCTACCACTGCCCCAGC 570
QY 3663 CGTGCTGGGAGCTGAAACCCACTTGGACTCTGGCCCTGCGCCACAGTGCATCTCTCCCA 3722
DB 571 CGTGCTGGGAGCTGAAACCCACTTGGACTCTGGCCCTGCGCCACAGTGCATCTCTCCCA 630
QY 3723 AGATGACGACTTCTGGCTGTGAGGAGCTGCAGACCTGGCTGTGGCGCTGGCCCAAG 3782
DB 631 AGATGACGACTTCTGGCTGTGAGGAGCTGCAGACCTGGCTGTGGCGCTGGCCCAAG 690
QY 3783 ACTTCAACCGGCTCAAGAAGATGAGCCCTCAGAGCTGACCTGACCTGACCTGACCTGG 3842
DB 691 ACTTCAACCGGCTCAAGAAGATGAGCCCTCAGAGCTGACCTGACCTGACCTGACCTGG 750
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTTCCTTCCCTCCCTTCCCT 3887
DB 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTTCCTTCCCTTCCCTTCCCT 795
RESULT 13
AX392086
LOCUS AX392086
DEFINITION Sequence 1 from Patent WO0215977.
ACCESSION AX392086
VERSION AX392086.1 GI:19700574
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Senaldi, G.
TITLE Methods and compositions for treating ige-related disease using nn
JOURNAL t-1 inhibitors
PATENT: WO 0215977-A 1 28-FEB-2002;
Amgen Inc. (US)
FEATURES Location/Qualifiers

Sun Feb 2 08:32:00 2003

us-09-931-704-3.rge

AY049779
 LOCUS AY049779 680 bp mRNA linear PRI 23-OCT-2001
 DEFINITION Homo sapiens cardiostrophin-like cytokine (CLC) mRNA, complete cds.
 ACCESSION AY049779
 VERSION AY049779.1 GI:16356642
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 680)
 Hu.X., Xu.Y., Zhang.B., Peng.X., Yuan.J. and Qiang.B.
 Direct Submission
 Submitted (30-JUL-2001) Department of Biochemistry, Institute of
 Basic Medical Science, Chinese Academy of Medical Sciences, 5 Dong
 Dan San Tiao, Beijing 100005, P.R. China
 Location/Qualifiers
 1..680
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..680
 /gene="CLC"
 3..680
 /gene="CLC"
 /note="neurotrophin-1/B-cell stimulating factor-3"
 /codon_start=1
 /product="cardiostrophin-like cytokine"
 /protein_id="AA15436.1"
 /db_xref="GI:16356643"
 /translations="MDLRAGDSWGLACLCTVLWHLPAVPAALNRTGDPGSPISQKTY
 DLTRYLEHQLRSAGTYLVNLYGPPNEPDPNPRPGAETLPRATVDLEWVRSNDKLR
 LTONYRAYSHLLCYLRLNQAAAEALRRSLAHECTSLQGLGSIAGVMAALGYPLPQ
 PLFGTEPTWTPGPAHSDFLQMDDFLWKLKELQTLWRSADFNRLKKMQPPAAAVIL
 HLGAGHF"

BASE COUNT 125 a 239 c 191 g 125 t
 ORIGIN

Query Match 9.7%; Score 495.4; DB 9; Length 680;
 Best Local Similarity 99.8%; Pred. No. 5.4e-110;
 Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCCCTGGGGG 3422
 DB 184 ATCTGAACCTACCTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCCCTGGGGG 243
 QY 3423 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAGCCCTCAATGACAAAC 3482
 DB 244 CAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAGCCCTCAATGACAAAC 303
 QY 3483 TGGGCTGACCCAGAACTACAGAGGCTACAGCCACCTTCTGTGTTACTTGGCTGGCTCA 3542
 DB 304 TGGGCTGACCCAGAACTACAGAGGCTACAGCCACCTTCTGTGTTACTTGGCTGGCTCA 363
 QY 3543 ACCGTGAGGCTGCCACTGTGAGCTGCGCGCAGCCCTGGGCCACTTCTGCACAGGCTCC 3602
 DB 364 ACCGTGAGGCTGCCACTGTGAGCTGCGCGCAGCCCTGGGCCACTTCTGCACAGGCTCC 423
 QY 3603 AGGGCTGTGGGAGCATTGGGGGGTGTATGGCAGCTCTGGGCTACCCACTGCCCCCAGC 3662
 DB 424 AGGGCTGTGGGAGCATTGGGGGGTGTATGGCAGCTCTGGGCTACCCACTGCCCCCAGC 483
 QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCCAAGTGAATCTCTCCAGA 3722
 DB 484 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCCAAGTGAATCTCTCCAGA 543
 QY 3723 AGATGACGACTTCTGGCTGCTGAAGAGCTGAGAGCTGGCTGTGGCTCGGCCAAGG 3782
 DB 544 AGATGACGACTTCTGGCTGCTGAAGAGCTGAGAGCTGGCTGTGGCTCGGCCAAGG 603
 QY 3783 ACTTCAACCGGCTCAAGAGAAAGATGAGCTCCAGAGCTGAGTCAACCTGACCTGG 3842
 DB 604 ACTTCAACCGGCTCAAGAGAAAGATGAGCTCCAGAGCTGAGTCAACCTGACCTGG 663

QY 3843 GGGCTCATGCTTCTGA 3859
 DB 664 GGGCTCATGCTTCTGA 680
 Search completed: February 1, 2003, 18:01:12
 Job time : 14154.7 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 10332.6 Seconds
(without alignments)
7973.437 Million cell updates/sec

Title: US-09-931-704-3
Perfect score: 5087
Sequence: 1 aacctgcagtgaggctggc.....ccttgcgaagtctctcca 5087

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estrio.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: gb_gss.*

18: em_gss_hum.*

19: em_gss_inv.*

20: em_gss_pln.*

21: em_gss_vrt.*

22: em_gss_fun.*

23: em_gss_mam.*

24: em_gss_mus.*

25: em_gss_other.*

26: em_gss_pro.*

27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860.2	16.9	955	9	AL543945
C 2	859.8	16.9	888	9	AL570325
C 3	656.6	12.9	681	10	AW978561
4	612	12.0	691	12	BF213570
5	567.4	11.2	915	12	BF035982
6	512.4	10.1	514	12	BF058969

7	488.6	9.6	1157	14	BQ940483
C 8	487	9.6	487	10	BE463575
C 9	479.4	9.4	493	9	AI040033
10	471.4	9.3	853	12	BG437538
11	470.4	9.2	887	14	BQ948158
C 12	465.4	9.1	467	12	BF002167
C 13	452.4	8.9	454	12	BF513712
C 14	450.4	8.9	457	10	AW044541
C 15	446.2	8.8	462	9	AI916548
C 16	429.4	8.6	465	9	AA430602
C 17	429.4	8.4	432	9	AA430517
C 18	422.4	8.3	432	9	AA278936
C 19	422.2	8.3	433	9	AI765382
C 20	407.4	8.0	409	9	AA283037
C 21	405	8.0	411	9	AI752562
C 22	399	7.8	407	9	AI752561
C 23	383.8	7.5	1053	12	BG164929
C 24	375	7.4	381	12	BF056328
C 25	374.2	7.4	452	9	AA283056
C 26	373.8	7.3	401	12	BF881796
C 27	372.6	7.3	396	12	BF849577
C 28	355.8	7.0	360	9	AA284882
C 29	352.4	6.9	354	9	AA830376
C 30	351	6.9	471	12	BF852246
C 31	350.4	6.9	594	14	BM763333
C 32	346.8	6.8	512	9	AI390475
C 33	343.2	6.7	356	9	AI769263
C 34	342.2	6.7	476	10	BE632644
C 35	342.2	6.6	622	12	BF852748
C 36	337.4	6.6	580	14	BM848189
C 37	333.2	6.6	448	12	BF853441
C 38	331.6	6.5	497	12	BF378449
C 39	331.4	6.5	573	14	BM840863
C 40	320	6.3	321	9	AI474292
C 41	319.4	6.3	346	9	AI868273
C 42	319.2	6.3	458	13	BM363136
C 43	316.4	6.2	351	9	AI571829
C 44	311.6	6.1	468	12	BF369558
C 45	311.2	6.1	457	12	BF750407

ALIGNMENTS

RESULT 1
AL543945
LOCUS
DEFINITION
AL543945 LTI_NFL006_PL2 Homo sapiens cdna clone CS0DI004YM15 5
prime, mRNA sequence.
ACCESSION
AL543945
VERSION
AL543945.1 GI:12876424
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 955)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI004YM15"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com

BASE COUNT 168 a 323 c 286 g 198 t
ORIGIN
Query Match 16.9%; Score 860.2; DB 9; Length 955;
Best Local Similarity 99.2%; Pred. No. 6.4e-186;
Matches 896; Conservative 0; Mismatches 3; Indels 4; Gaps 3;
QY 3363 AGCTGAACCTACCTGGCCGCCCTTTCAACGAGCGACACTTCAACCCCTCCCGCTGGGG 3422
Db 53 ATCTGAACCTACCTGGCCGCCCTTTCAACGAGCGACACTTCAACCCCTCCCGCTGGGG 112
QY 3423 CAGAGACTTCCCGAGGCGCACTTTGACTTGGAGGTGCGGAAGCTCAATGACAAAC 3482
Db 113 CAGAGACTTCCCGAGGCGCACTTTGACTTGGAGGTGCGGA--CTCAATGACAAAC 170
QY 3483 TCGCGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTGTACTTGGTGGCTCA 3542
Db 171 TCGCGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTGTACTTGGTGGCTCA 230
QY 3543 ACCGTGAGGTGCACTGTGAGTGGCGCGAGCTGGCCACTTCTGCACGAGCTCC 3602
Db 231 ACCGTGAGGTGCACTGTGAGTGGCGCGAGCTGGCCACTTCTGCACGAGCTCC 290
QY 3603 AGGGCTGCTGGCGAGCACTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAGC 3662
Db 291 AGGGCTGCTGGCGAGCACTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCAGC 350
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCCAGA 410
Db 351 CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTTCTCCAGA 410
QY 3723 AGATGAGCACTTCTGCTGCTGAGAGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 3782
Db 411 AGATGAGCACTTCTGCTGCTGAGAGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
QY 3783 ACTTCAACCGGCTCAAGAGAGATGAGAGCTCCAGAGCTGCACTGCACTGCACTGCTG 3842
Db 471 ACTTCAACCGGCTCAAGAGAGATGAGAGCTCCAGAGCTGCACTGCACTGCACTGCTG 530
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 3902
Db 531 GGGCTCATGGCTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 590
QY 3903 CACTTTGTGAGAGCGAGCCCTGTATGCCAACACCTGTTGAGCGAGGACAGAACTGTG 3962
Db 591 CACTTTGTGAGAGCGAGCCCTGTATGCCAACACCTGTTGAGCGAGGACAGAACTGTG 650
QY 3963 AGCCTCTGGCCCTTTCTCTGGACCGGCTGGGGCTGTGATCGATCAGCCCTGTCTCTCC 4022
Db 651 AGCCTCTGGCCCTTTCTCTGGACCGGCTGGGGCTGTGATCGATCAGCCCTGTCTCTCC 710
QY 4023 CACTTCCCAAGAGTCTACCGAGCT--GGGAGGAGGTACAGTAGGCGCTGCTGCTGCTG 4081
Db 711 CACTTCCCAAGAGTCTACCGAGCT--GGGAGGAGGTACAGTAGGCGCTGCTGCTGCTG 770
QY 4082 TTCTACAGGAAGTCAATGCTCGAGGAGGTGTGAAGTGGTTCAGGTGGTGGTGCAGAGGCGCTC 4141
Db 771 TTCTACAGGAAGTCAATGCTCGAGGAGGTGTGAAGTGGTTCAGGTGGTGGTGCAGAGGCGCTC 830
QY 4142 ATGGCTCCTGCTTCTGCTTACCACTTGGCCAGTGGCCCAACCCAGCCCTCAGGTGGCAC 4201
Db 831 ATGGCTCCTGCTTCTGCTTACCACTTGGCCAGTGGCCCAACCCAGCCCTCAGGTGGCAC 890
QY 4202 ATCTGGAGGCA--GGGTTGAGGGGCCACACACATGCTTCTCTGGGTTGAAGCCCT 4260

Db 891 ATCTGAGGCGAGGCGGTTGAGGGGCCACACACATGCTTCTGGGTTGAAGCCTT 950
QY 4261 TTG 4263
Db 951 TGG 953
RESULT 2
AL570325/c 888 bp mRNA linear EST 16-FEB-2001
LOCUS AL570325 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1004YM15 3
DEFINITION prime, mRNA sequence.
ACCESSION AL570325
VERSION AL570325.1 GI:12926520
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 888)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clones="CS0D1004YM15"
/tissue_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 204 a 233 c 247 g 199 t 5 others
ORIGIN
Query Match 16.9%; Score 859.8; DB 9; Length 888;
Best Local Similarity 99.0%; Pred. No. 7.8e-186;
Matches 880; Conservative 5; Mismatches 2; Indels 2; Gaps 2;
QY 3884 CCCCTTCAACCCCTGCTCCCACTTTGTGAGAGCGAGCCCTGTATG--CCAACACCTGTGA 3942
Db 888 CCCCTTCAACCCCTGCTCCCACTTTGTGAGAGCGAGCCCTGTATGCTTACCTGTGA 829
QY 3943 GCAGAGACAGAACTGTGAGCTCTGGCCCTTCTCTGACCGCTGGCGTGTGATGC 4002
Db 828 GCAGAGACAGAACTGTGAGCTTTTGGCCYTTTCTGACCGCTGGCGTGTGATGC 769
QY 4003 GATCAGCCCTGTCTCTCCCACTCCCAAGGTTTACCGAGCTGGGGAGAGGTACAGT 4062
Db 768 GATCAGCCCTGTCTCTCCCACTCCCAAGGTTTACCGAGCTGGGGAGAGGTACAGT 709
QY 4063 AGGCCCTGTCTCTCTGTTTCTACAGGAAGTATGCTCAGGAGGTGTGAAGTGTTC 4122
Db 708 AGGCCCTGTCTCTCTGTTTCTACAGGAAGTATGCTCAGGAGGTGTGAAGTGTTC 649
QY 4123 GGTGTGTGACAGCGCTCATGCTCTCTGCTTCTTCCCTACCACTTGGCCAGTGGCCAC 4182
Db 648 GGTGTGTGACAGCGCTCATGCTCTCTGCTTCTTCCCTACCACTTGGCCAGTGGCCAC 589
QY 4183 CCAGCCCTCAGTGGGCATATCTGAGGAGCGAGGGTTGAGGGGCCACACACATGCC 4242

```
Db 588 .CAGCCCTCAGGTGGACATCTCGAGGGGAGGGGTTGAGGGGCCACACACATGCC 529
QY 4243 TTTCTGGGTGAAGCCCTTTGGCTGCCCACTCTCTTGGATGGTGTGCTCCCTTATC 4302
Db 528 TTTCTGGGTGAAGCCCTTTGGCTGCCCACTCTCTTGGATGGTGTGCTCCCTTATC 469
QY 4303 CCGAAATCACTTATACATCAATTCAGGAACAAACATGCTGGAATTTACACAAAA 4362
Db 468 CCCAAATCACTTATACATCAATTCAGGAACAAACATGCTGGAATTTACAC-AAAA 410
QY 4363 GAGATGAGATTAAACAGTGCAGGGTTGGGCTCTGATTGGAGTGCCCTATATAACACAGAAG 4422
Db 409 GAGATGAGATTAAACAGTGCAGGGTTGGGCTCTGATTGGAGTGCCCTATATAACACAGAAG 350
QY 4423 AGAAATACCTGAAGACAGAGGGGAGGAGACAGACAGACAGACAGAGGAGTCTCCAAAG 4482
Db 349 AGAAATACCTGAAGACAGAGGGGAGGAGACAGACAGACAGACAGAGGAGTCTCCAAAG 290
QY 4483 CACAGATGGCAACAAACCCGAGCTGAGCATCAGACCTTGCTCGAATTTGCTTCCA 4542
Db 289 CACAGATGGCAACAAACCCGAGCTGAGCATCAGACCTTGCTCGAATTTGCTTCCA 230
QY 4543 GTATTACGCTGCTTCTCTGCCCCCTTTCCAGGGTATCTGCGGTGCGAGGCTGGG 4602
Db 229 GTATTACGCTGCTTCTCTGCCCCCTTTCCAGGGTATCTGCGGTGCGAGGCTGGG 170
QY 4603 GAGGGCAACCATAGCCACACACAGGATTTCTCTGAAAGTTTACAAATGCAAGTATTTG 4662
Db 169 GAGGGCAACCATAGCCACACACAGGATTTCTCTGAAAGTTTACAAATGCAAGTATTTG 110
QY 4663 GGGTGTAGGTGGAGCTCCCAAGGCCCTGCCCCCAGCCCAACCCCACTCATGACTCTA 4722
Db 109 GGGTGTAGGTGGAGCTCCCAAGGCCCTGCCCCCAGCCCAACCCCACTCATGACTCTA 50
QY 4723 AGTGTGTTGATTAAATTTATTATTATTGGAGATGTTATTATTAGATG 4771
Db 49 AGMTGTTGATTAAATTTATTATTATTGGAGATGTTATTATTAGATG 1

RESULT 3
AW978561/c 681 bp mRNA linear EST 02-JUN-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
I-E, Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL
Unpublished (2000)
COMMENT
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 394
Seq primer: Forward.
Location/Qualifiers
1. .681
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone_lib="WAGE resequences, MAGP"
/note="Vector: pBluescriptSKm"

BASE COUNT 153 a 94 c 248 g 185 t 1 others
ORIGIN
```

```
Query Match 12.9%; Score 656.6; DB 10; Length 681;
Best Local Similarity 99.3%; Pred. No. 2.3e-139;
Matches 670; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 190 TGATACCTAAACGACCAAGTCACAGCCCTCCAACTCACCTCTGCGCTGCCACACCTCA 249
Db 681 TAAATACCTAAACGACCAAGTCACAGCCCTCCAACTCACCTCTGCGCTGCCACACCTCA 622
QY 250 CCACATCCTTGG-TGGAGCTCAAACTCAACGACCTAAATCAACCAATCCCAAGTCTAAA 308
Db 621 CCACATCCTTGGTGGAGCTCAAACTCAACGACCTAAATCAACCAATCCCAAGTCTAAA 562
QY 309 CTAATCTGAACCTTTTAAAGTAAACCCAGTCTCTTAAACCTTAACCTAGGCCAATGCAATTA 368
Db 561 CTAATCTGAACCTTTTAAAGTAAACCCAGTCTCTTAAACCTTAACCTAGGCCAATGCAATTA 502
QY 369 TATCTACCTAGCCAAACCCCTTAACCTGCTTTCAGTCCAGTCCAAAGTGTCCACATTCCTCA 428
Db 501 TATCTACCTAGCCAAACCCCTTAACCTGCTTTCAGTCCAGTCCAAAGTGTCCACATTCCTCA 442
QY 429 CTTTGGTCTCCTCAGTGAATAATCCCAAGAAAGCATATTTCCCACTGCCACATCCCTCCTT 488
Db 441 CTTTGGTCTCCTCAGTGAATAATCCCAAGAAAGCATATTTCCCACTGCCACATCCCTCCTT 382
QY 489 ACAGCACCCCAACCCCTGCGCTCTGGACTCTCTGGTATCTCTGGGATGTCCTCAAACTCTGCGATG 548
Db 381 ACAGCACCCCAACCCCTGCGCTCTGGACTCTCTGGTATCTCTGGGATGTCCTCAAACTCTGCGATG 322
QY 549 CCATCAGCACAAGAGCCGAGTCTGTAATGACCTCTCTCCCTTCTGTCCTCCCACTCCCTT 608
Db 321 CCATCAGCACAAGAGCCGAGTCTGTAATGACCTCTCTCCCTTCTGTCCTCCCACTCCCTT 262
QY 609 GCAGGCTGATGGAAGAGCCCTCATTTGAAGTCCAACTTTTCCCACTAACACCAAGAACGG 668
Db 261 GCAGGCTGATGGAAGAGCCCTCATTTGAAGTCCAACTTTTCCCACTAACACCAAGAACGG 202
QY 669 GGTGAACCTCCACACTGCCACCTGCTTCCCTGAGAGTGAGCACTAAATCTCTCTCAATCTAA 728
Db 201 GGTGAACCTCCACACTGCCACCTGCTTCCCTGAGAGTGAGCACTAAATCTCTCTCAATCTAA 142
QY 729 CCCACCTCTACACTTCCCACTCAGGAATACATCTCTAGATATATACCAAACTAAGCC 788
Db 141 CCCACCTCTACACTTCCCACTCAGGAATACATCTCTAGATATATACCAAACTAAGCC 82
QY 789 CCATAAGGACAGCCGAGCCCTAGTGGTCTTAACCTTATACCTTCTCTATGGGTGAGTCT 848
Db 81 CCATAAGGACAGCCGAGCCCTAGTGGTCTTAACCTTATACCTTCTCTATGGGTGAGTCT 22
QY 849 GTTCTTGGCGGCGCG 863
Db 21 GTTCTTGGCGGCGCG 7

RESULT 4
BF213570 691 bp mRNA linear EST 06-NOV-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 691)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
```

cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM914 row: 1 column: 17
High quality sequence stop: 637.

FEATURES

Location/Qualifiers
1. .691
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4070560"
/clone_lib="NIH MGC 55"
/tissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGGAGGCGGCACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 141 a 203 c 192 g 155 t
ORIGIN
Query Match 12.0%; Score 612; DB 12; Length 691;
Best Local Similarity 94.5%; Pred. No. 3.7e-129;
Matches 656; Conservative 0; Mismatches 35; Indels 3; Gaps 2;
Qy 3727 GGACGACTTCTGGCTCTGAAGAGCTGCAGACCTGGCTGTGGCGTCGCGCCAAAGACTT 3786
Db 1 GGACGACTTCTGGCTCTGAAGAGCTGCAGACCTGGCTGTGGCGTCGCGCCAAAGACTT 60
Qy 3787 CAACGGCTCAGAGAGATGACGCTCCAGAGCTGCAGTCAACCTGCACCTGGGGG 3846
Db 61 CAACGGCTCAGAGAGATGACGCTCCAGAGCTGCAGTCAACCTGCACCTGGGGG 120
Qy 3847 TCATGGCTTCTGACCTTCTCCTCTTCTGCTCCCTCTTCAAAACCTGCTCCACT 3906
Db 121 TCATGGCTTCTGACCTTCTCCTCTTCTGCTCCCTCTTCAAAACCTGCTCCACT 180
Qy 3907 TTGTGAGAGCCAGCCTGTATGCCAACACCTGTTGAGCAGGAGACAGAGCTGTGAGCC 3966
Db 181 TTGTGAGAGCCAGCCTGTATGCCAACACCTGTTGAGCAGGAGACAGAGCTGTGAGCC 240
Qy 3967 TCTGGCCCTTCTGACCGCTGGCGGTGTATGCGATCAGCCCTGCTCTCCCTCCACC 4026
Db 241 TCTGGCCCTTCTGACCGCTGGCGGTGTATGCGATCAGCCCTGCTCTCCCTCCACC 300
Qy 4027 TCCCAAGGTCTACGAGCTGGGAGAGGTACAGTAGGCCCTGTCTGCTCTGTTCTA 4086
Db 301 TCCCAAGGTCTACGAGCTGGGAGAGGTACAGTAGGCCCTGTCTGCTCTGTTCTA 360
Qy 4087 CAGGAAGTCATGCTCAGGGAGTGAAGTGGTTTCAGGTTGGTGCAGAGCGCTCATGGC 4146
Db 361 CAGGAAGTCATGCTCAGGGAGTGAAGTGGTTTCAGGTTGGTGCAGAGCGCTCATGGC 420
Qy 4147 CTCCTGCTTCTTCTTACCATTTGGCAGTGGCCACCCAGCCCTCAGGTGGCAGCATCTG 4206
Db 421 CTCCTGCTTCTTCTTACCATTTGGCAGTGGCCACCCAGCCCTCAGGTGGCAGCATCTG 480
Qy 4207 GAGGGCAGGGGTGAGGGGCGCCACACATGCTCTTCTGCGGTGAAGCCCTTTGGCT 4266
Db 481 GAGGGCAGGGGTGAGGGGCGCCACCA-CAAAATGCTCTTCTGCGGTGAAGCCCTTTGG- 537
Qy 4267 GCCCAGCTCTCTTGAATGGGTGCTGCTCTTATCCCAAAATCACTCTATATCAATCAAT 4326
Db 538 TGCCCACTCTCTTGAATGGGTGCTGCTCTTATCCCAAAATCACTCTATATCAATCAAT 597

Qy 4327 TCAGGAAACAAACATGTTGGCAATTTTACACAAAAGAGATGAGATTAAACAGTCAGGGT 4386
Db 598 TCAGGAAACAAACATGTTGGCAATTTTACACAAAAGAGATGAGATTAAACAGTCAGGC 657
Qy 4387 TGGGGTCTGCAATGGAGGTGCGCCCTATTAACACAGA 4420
Db 658 TGGGGTCTGATTTGGAGGGGCTATAACACAGAAGA 691

RESULT 5

BF035982 915 bp mRNA linear EST 20-OCT-2000
LOCUS 601457860F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3861542 5',
DEFINITION mRNA sequence.
ACCESSION BF035982
VERSION BF035982.1 GI:10743697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM9598 row: 9 column: 15
High quality sequence stop: 615.

FEATURES

Location/Qualifiers
1. .915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3861542"
/clone_lib="NIH MGC 66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: Sall; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 184 a 285 c 250 g 196 t
ORIGIN

Query Match 11.2%; Score 567.4; DB 12; Length 915;
Best Local Similarity 87.7%; Pred. No. 6.2e-119;
Matches 701; Conservative 0; Mismatches 86; Indels 12; Gaps 7;

Qy 3638 GCTCTGGCTACCCACCTGCCCGCCGCTGCTGGAGCTGAACCCACTTGGACTCTGTGC 3697
Db 6 GCTCTGGCTACCCACCTGCCCGCCGCTGCTGGAGCTGAACCCACTTGGACTCTGTGC- 64
Qy 3698 CCTGCCACAGTGAATTCCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGCTGCAG 3757
Db 65 CCTGCCACAGTGAATTCCTCCAGAGATGACGACTTCTGGCTGCTGAAGAGCTGCAG 124
Qy 3758 ACCTGGCTGTGGCGCTCGGCCCAAGGATTCAACCGGCTCAAGAGAGATGAGCGCTCCA 3817
Db 125 ACCTGGCTGTGGCGCTCGG-CAAGGACTTCAACCGGCTCAAGAGAGATGAGCGCTCCA 183
Qy 3818 GCGAGCTGAGTCAACCTGCGACTCGGGGCTCATGCTTCTGACTTCTGACCTTCTCTCT 3877
Db 184 GCGAGCTGAGTCAACCTGCGACTCGGGGCTCATGCTTCTGACTTCTGACCTTCTCTCT 243
Qy 3878 TCGCTCCCTTCAACCCCTCTCCCACTTTTGTGAGAGCCAGCCCTGTATGCCAACACCT 3937

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13608 row: g column: 03
 High quality sequence stop: 572.
 Location/Qualifiers
 1. 1157
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6197786"
 /clone_lib="Lupski_sciatic_nerve"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
 Not1; Site_2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCAGCGTCGCG-3' and
 5'-GACTAGTCTAGATCGGCGCGCCCT(15)-3'. Size selected >
 1 kb for average insert length 1.87 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life
 Technologies."

BASE COUNT 221 a 374 c 346 g 216 t
 ORIGIN

Query Match 9.6%; Score 488.6; DB 14; Length 1157;
 Best Local Similarity 94.6%; Pred. No. 6.8e-101;
 Matches 529; Conservative 0; Mismatches 24; Indels 6; Gaps 2;
 QY 3363 AGCTGAACCTACCTGGGCGGCGGCTTCAACGAGCCAGACTTCAACCTCCCGCGCTGGGG 3422
 DB 225 ATCTGAACCTACCTGGGCGGCGGCTTCAACGAGCCAGACTTCAACCTCCCGCGCTGGGG 284
 QY 3423 CAGAGACTCTGCCAGGCGGCACTGTTGACTTTGGAGGTGTGGCGAGGCTCAATGACAAAC 3482
 DB 285 CAGAGACTCTGCCAGGCGGCACTGTTGACTTTGGAGGTGTGGCGAGGCTCAATGACAAAC 344
 QY 3483 TGGCGCTGACCCAGAACTAGAGCGCTACAGCGCTTCTGTGTACTTGTGGTGGCTCA 3542
 DB 345 TGGCGCTGACCCAGAACTAGAGCGCTACAGCGCTTCTGTGTACTTGTGGTGGCTCA 404
 QY 3543 ACCGTGAGGCTGCCACCTGCTGAGCTGCGCGCAGCTGGCGCACTTCTGACCGAGCTCC 3602
 DB 405 ACCGTGAGGCTGCCACCTGCTGAGCTGCGCGCAGCTGGCGCACTTCTGACCGAGCTCC 464
 QY 3603 AGGCGCTGCTGGGCGAGCACTTGGGCGGCTCATGGCAGCTCTGGGCTACCACTGGCCCGAGC 3662
 DB 465 AGGCGCTGCTGGGCGAGCACTTGGGCGGCTCATGGCAGCTCTGGGCTACCACTGGCCCGAGC 524
 QY 3663 CGTGCCTGGGACTGAACCCACTTGGACTCTGGCGCTGCCCGCAGCTGACTTCTCCAG 3722
 DB 525 CGTGCCTGGGACTGAACCCACTTGGACTCTGGCGCTGCCCGCAGCTGACTTCTCCAG 584
 QY 3723 AGATGAGCAGCTTCTGGCTGTGAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG 3782
 DB 585 AGATGAGCAGCTTCTGGCTGTGAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG 644
 QY 3783 ACTTCAACCGGCTCAAGAAGAGATGACGCTTCCAGCGTGCAGTCAACCTGCACTGG 3842
 DB 645 ACTTCAACCGGCTCAAGAAGAGATGACGCTTCCAGCGTGCAGTCAACCTGCACTGG 704
 QY 3843 GGG--CTCATGGCTTCTGACTTCTGAC---CTTCTCTCTCTGCTCCCGCTTCAAAACC 3896
 DB 705 GGGCTCATGGGCTTCTGACTTCTGACCTTCTCTCTCTCTGCTCCCGCTTCAAAACC 764

QY 3897 TGCTCCCACTTTGTGAGAG 3915

DB 765 TGCTCCCACTTTGTGAG 783

RESULT 8

BE463575/c

LOCUS BE463575 487 bp mRNA linear EST 27-JUL-2000
 DEFINITION hw25d05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3183945 3',
 mRNA sequence.

ACCESSION BE463575

VERSION BE463575.1 GI:9509350

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 487)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40Up from Gibco

High quality sequence stop: 452.

Location/Qualifiers

1. 487

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3183945"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneids 132376-132391, 1456007-1456775, and

1509552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo. "

BASE COUNT 122 a 117 c 114 g 134 t

ORIGIN

Query Match 9.6%; Score 487; DB 10; Length 487;
 Best Local Similarity 100.0%; Pred. No. 1.4e-100;
 Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4339 CATGTGGCAATCTACACAAAAGAGATGAGATTAAACAGTGCAGGGTTCGGGTCTGCAT 4398

DB 487 CATGTGGCAATCTACACAAAAGAGATGAGATTAAACAGTGCAGGGTTCGGGTCTGCAT 428

QY 4399 TGGAGGTGCCCTATAAACACAGAGAAATACTGAAAGCACAGGGGCGAGGACAGACCA 4458

DB 427 TGGAGGTGCCCTATAAACACAGAGAAATACTGAAAGCACAGGGGCGAGGACAGACCA 368

QY 4459 GACACAGCCAGGAGTCTCCAAAGCAGAGTGCAGCAAAACCCAGCTGACATCAG 4518

DB 367 GACACAGCCAGGAGTCTCCAAAGCAGAGTGCAGCAAAACCCAGCTGACATCAG 308

QY 4519 GACCTTGCCTCGAATGTCTTCAGTATTACGGTGCCTCTTCTTGCCTTTCACAGG 4578

```

Db 307 GACCTTGCTCGAATTGCTCCAGATTACGGTGCCTCTTCTGCCCCCTTTCCCGAGG 248
Qy 4579 GTATCTGGGTGTCAGGCTGGGAGGGCAACCATAGCCACACACAGGATTTCTCGAA 4638
Db 247 GTATCTGGGTGTCAGGCTGGGAGGGCAACCATAGCCACACACAGGATTTCTCGAA 188
Qy 4639 AGTTTACAATGAGTAGATTTTGGGGTGTAGGGTGGGAGCTCCCAAGGGCCTGCCCCC 4698
Db 187 AGTTTACAATGAGTAGATTTTGGGGTGTAGGGTGGGAGCTCCCAAGGGCCTGCCCCC 128
Qy 4699 CAGCCCCACCCACTCATGACTCTAAGTGTGTGTATTATTAATTTATTTATTTGGAGATGT 4758
Db 127 CAGCCCCACCCACTCATGACTCTAAGTGTGTGTATTATTAATTTATTTATTTGGAGATGT 68
Qy 4759 TATTTATTAGATGATTTATTATTCAGAAATTTCTATTTCTGATTAAACAAATAAAATGCTT 4818
Db 67 TATTTATTAGATGATTTATTATTCAGAAATTTCTATTTCTGATTAAACAAATAAAATGCTT 8
Qy 4819 GCCCGAG 4825
Db 7 GCCCGAG 1

RESULT 9
AI040033/c
LOCUS 493 bp mRNA linear EST 24-SEP-1998
DEFINITION ox97h03.x1 Soares senescent fibroblasts_NbHSF Homo sapiens CDNA
clone IMAGE:1664309 3', mRNA sequence.
ACCESSION AI040033
VERSION AI040033.1 GI:3279227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 954 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 456.
FEATURES
Location/Qualifiers
1..493
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:1664309"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent_fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/notes="vector: p773D (Pharmacia) with a modified
polylinker V TYPE: phagemid; Site 1: Not I; Site 2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTACCACTCTGAAGTGGGAGGGCGGCGGATTTTTTTTTTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 121 a 113 c 112 g 147 t
ORIGIN

Query Match 9.4%; Score 479.4; DB 9; Length 493;
Best Local Similarity 99.8%; Pred. No. 7.6e-99;
Matches 480; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 4347 CAATTTACACAAAAGAGATGAGATTAACTAGTCAGGGTTGGGCTCTGCATTGGAGGTG 4406
Db 493 CAATTTACACAAAAGAGATGAGATTAACTAGTCAGGGTTGGGCTCTGCATTGGAGGTG 434
Qy 4407 CCCTATAACACAGAAGAAAATCTGAAAGCAGAGGGGAGGGACAGACCAAGACAGAC 4466
Db 433 CCCTATAACACAGAAGAAAATCTGAAAGCAGAGGGGAGGGACAGACCAAGACAGAC 374
Qy 4467 CCAGAGGTCTCCAAAGCAGAGTGGCAACAAAACCCGAGCTGAGCATCAGGACCTTGC 4526
Db 373 CCAGAGGTCTCCAAAGCAGAGTGGCAACAAAACCCGAGTTCAGCATCAGGACCTTGC 314
Qy 4527 CTCGAATTTGCTTTCCAGTATTACGGTGCCTCTTCTGCCCCCTTTCCAGGGTATCTGT 4586
Db 313 CTCGAATTTGCTTTCCAGTATTACGGTGCCTCTTCTGCCCCCTTTCCAGGGTATCTGT 254
Qy 4587 GGGTTGCCAGGCTGGGAGGGCAACCATAGCCACACACAGGATTTCTGAAAGTTTACA 4646
Db 253 GGGTTGCCAGGCTGGGAGGGCAACCATAGCCACACACAGGATTTCTGAAAGTTTACA 194
Qy 4647 ATGCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCAAGGCCCTGCCGCCAGCCCA 4706
Db 193 ATGCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCAAGGCCCTGCCGCCAGCCCA 134
Qy 4707 CCCACTCATGACTCTAAAGTGTGTTGTTATTAATTTATTTATTTGGAGATGTTATTTATT 4766
Db 133 CCCACTCATGACTCTAAAGTGTGTTGTTATTAATTTATTTATTTGGAGATGTTATTTATT 74
Qy 4767 AGATGATATTTATTGCGAATTTCTATTCTTGTTATTAACAAATAAAATGCTTGCCTCC 4826
Db 73 AGATGATATTTATTGCGAATTTCTATTCTTGTTATTAACAAATAAAATGCTTGCCTCC 14
Qy 4827 A 4827
Db 13 A 13

RESULT 10
BG437538
LOCUS 853 bp mRNA linear EST 14-MAR-2001
DEFINITION 602489277F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4621530 5',
mRNA sequence.
ACCESSION BG437538
VERSION BG437538.1 GI:13344044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTG/Gazdar
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI384 row: i column: 19
High quality sequence stop: 761.
FEATURES
Location/Qualifiers
1..853
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:4621530"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:

```



```

RESULT 12.
BF002167/c
LOCUS
DEFINITION
7h23c10.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:3316818 3',
mRNA sequence.
ACCESSION
BF002167
VERSION
BF002167.1 GI:10702442
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 467)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
FEATURES
source
1..467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3316818"
/clone_lib="NCI_CGAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 117 a 113 c 108 g 129 t
ORIGIN
Query Match 9.1%; Score 465.4; DB 12; Length 467;
Best Local Similarity 99.8%; Pred. No. 1.2e-95;
Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4360 AAGAGATGAGATTAAACAGTGCAGGGTGGGTCTGCAATTGGAGTGCCCTATTAACACG 4419
DB 467 AAGAGATGAGATTAAACAGTGCAGGGTGGGTATGCAATTGGAGTGCCCTATTAACACG 408
QY 4420 AAGAGAAATACTGAAGACAGGGGCGAGACAGACAGACAGACCCAGGAGTCTCCA 4479
DB 407 AAGAGAAATACTGAAGACAGGGGCGAGACAGACAGACAGACCCAGGAGTCTCCA 348
QY 4480 AAGCACAGAGTGGCAAAACAAACCCGAGCTGAGCATCAGGACCTTGCTCGAATTGTCTT 4539
DB 347 AAGCACAGAGTGGCAAAACAAACCCGAGCTGAGCATCAGGACCTTGCTCGAATTGTCTT 288
QY 4540 CCAGTATTAGTGCCTCTTCTCTGCCCCCTTTCCAGGGTATCTGTGGGTGCCAGGCT 4599
DB 287 CCAGTATTAGTGCCTCTTCTCTGCCCCCTTTCCAGGGTATCTGTGGGTGCCAGGCT 228
QY 4600 GGGGAGGGCAACCATAGCCACACACAGGATTTCTGAAAGTTTACATGCAGTAGCATT 4659
DB 227 GGGGAGGGCAACCATAGCCACACACAGGATTTCTGAAAGTTTACATGCAGTAGCATT 168

```

```

QY 4660 TTGGGGTGTAGGTGGCAGCTCCCAAGGCCCTGCCCCCAGCCGCCACCTCATGACT 4719
DB 167 TTGGGGTGTAGGTGGCAGCTCCCAAGGCCCTGCCCCCAGCCGCCACCTCATGACT 108
QY 4720 CTAAGTGTGTGATTAAATATTATTATTGAGAGATGTTATTATTAGATGATATTAT 4779
DB 107 CTAAGTGTGTGATTAAATATTATTATTGAGAGATGTTATTATTAGATGATATTAT 48
QY 4780 TGCAGAATTTCTATTCTTGTATTACAAATAAAATGCTTCCCCCAGA 4826
DB 47 TGCAGAATTTCTATTCTTGTATTACAAATAAAATGCTTCCCCCAGA 1
RESULT 13
BF513712/c
LOCUS
DEFINITION
UI-H-BW1-amy-c-07-0-UI.sl NCI_CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3071604 3', mRNA sequence.
ACCESSION
BF513712
VERSION
BF513712.1 GI:11598891
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 454)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapbs-r@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=No.
FEATURES
source
1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3071604"
/clone_lib="NCI_CGAP_Sub7"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI_CGAP_Sub7
is a subtracted library derived from NCI_CGAP_Sub6. The
NCI_CGAP_Sub7 library had 12 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
132376-132391, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid3 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE CloneIDs 1414920-1417991, 1520904-152439
); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735
(IMAGE CloneIDs 1257096-1258631, 1469064-1470983
, 1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
1101192-1101959, 1217928-1220615); NCI_CGAP_Col0 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255
, 1144584-1145351). (6% of the driver population), plus a
pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE
CloneIDs 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE
CloneIDs 2710536-2712455) (4% of the driver population
), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE
CloneIDs 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
CloneIDs 2723592-2729326) (40% of the driver population),

```


REFERENCE 1 (bases 1 to 462)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
CDNA Library Arrayed by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 600 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 445.

FEATURES
source
1..462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2299261"
/clone_lib="NCI CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 114 a 108 c 108 g 132 t
ORIGIN

Query Match 8.8%; Score 446.2; DB 9; Length 462;
Best local Similarity 99.3%; Pred. No. 3e-91;
Matches 448; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4377 AGTCAGGGTTGGGCTCTGCATTGGAGTGGCCCTATAAACAGAGAGAAAACTGAAA 4436
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
462 AGTCAGGGTTGGGCTCTGCATTGGAGTGGCCCTATAAACAGAGAGAAAACTGAAA 403
QY 4437 GCACAGGGGAGGACAGACAGACAGACAGACAGAGTCTCCAAAGCACAGAGTGGCAA 4496
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
402 GCACAGGGGAGGACAGACAGACAGACAGACAGAGTCTCCAAAGCACAGAGTGGCAA 343
QY 4497 CAAAACCCGAGCTGAGCATCAGGACCTTGCCCTCGAAATTTGCTTCCAGTATTACGGTGCCT 4556
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
342 CAAAACCCGAGCTGAGCATCAGGACCTTGCCCTCGAAATTTGCTTCCAGTATTACGGTGCCT 283
QY 4557 CTTCTCTGCCCCCTTTCCAGGGTATCTGTGGGTTGTCAGGCTGGGGAGGGCAACCATAG 4616
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
282 CTTCTCTGCCCCCTTTCCAGGGTATCTGTGGGTTGTCAGGCTGGGGAGGGCAACCATAG 223
QY 4617 CCACACACAGAGATTTCTCTGAAAGTTTACATGAGTAGCATTTTGGGGTGTAGGGTGGC 4676
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
222 CCACACACAGAGATTTCTGAAAGTTTACATGAGTAGCATTTTGGGGTGTAGGGTGGC 163
QY 4677 AGCTCCCAAGGCCCTGCCCCCAGCCCCCACCCTCATGACTCTAAGTGTGTGTATTA 4736
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
162 AGCTCCCAAGGCCCTGCCCCCAGCCCCCACCCTCATGACTCTAAGTGTGTGTATTA 103
QY 4737 ATATTATTATTGGAGATGTTTATTATTAGATGATATTATTGCGAATTTCTATTCT 4796
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
102 ATATTATTATTGGAGATGTTTATTATTAGATGATATTATTGCGAATTTCTATTCT 43
QY 4797 TGTATTAAACAAATAAATGCTTGGCCAGAA 4827
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
42 TGTATTAAACAAATAAATGCTTGGCCAGAA 12

Search completed: February 1, 2003, 12:59:11
Job time : 10371.6 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1295.47 Seconds
(without alignments)
8843.084 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctggagtgccctggc.....ccttgtaagtctctcaca 5087

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq 101002.*

1: /SID22/gcgdata/geneq/geneq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneq/geneq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneq/geneq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneq/geneq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneq/geneq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneq/geneq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneq/geneq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneq/geneq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneq/geneq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneq/geneq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneq/geneq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneq/geneq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneq/geneq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneq/geneq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneq/geneq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneq/geneq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneq/geneq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneq/geneq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneq/geneq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneq/geneq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneq/geneq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneq/geneq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneq/geneq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5087	100.0	5087	21	AAA39482 Human NNT-1 DNA.
2	5087	100.0	5087	24	ABK11648 Human novel neurot
3	5076	99.8	5088	19	AAV47511 Human neurotrophic
4	5076	99.8	5088	19	AAV22653 Human genomic DNA
5	1464.4	28.8	1790	21	AAA88546 Human interleukin-
6	1463.4	28.8	1710	20	AAK16161 Human cardiophi
7	535.4	10.5	1008	22	AAK51548 Human polynucleoti
8	529.4	10.4	729	22	AAO4201 Human cardiophi
9	525.4	10.3	861	22	AAH74484 Nucleotide sequenc

10	523.4	10.3	797	19	AAV47510	Human neurotrophic
11	523.4	10.3	797	19	AAV22652	cDNA encoding huma
12	523.4	10.3	797	21	AAA39481	Human NNT-1 cDNA
13	523.4	10.3	797	24	ABK11647	Human cDNA encodin
14	485.4	9.7	968	22	ABA09140	Human cardiophi
15	495.4	9.7	968	22	AAK52532	Human polynucleoti
16	493.4	9.7	495	22	ABA71773	Human fetal liver
17	493.4	9.7	495	22	ABA37855	Probe #16321 for g
18	493.4	9.7	495	22	AAK20142	Human brain expres
19	493.4	9.7	495	22	AAK46202	Human bone marrow
20	493.4	9.7	495	22	AAI25564	Probe #15497 for g
21	493.4	9.7	495	22	AAI52108	Probe #20794 used
22	493.4	9.7	495	22	ABS20513	Human genome-deriv
23	492.4	9.7	768	22	AAH99772	Human protein enco
24	492.4	9.7	492	22	ABA59256	Human fetal liver
25	492	9.7	492	22	ABA27996	Probe #6462 for ge
26	492	9.7	492	22	AAK07469	Human brain expres
27	492	9.7	492	22	AAK33253	Human bone marrow
28	492	9.7	492	22	AAI16384	Probe #6317 for ge
29	492	9.7	492	22	AAI39047	Probe #7733 used t
30	492	9.7	492	22	ABS08085	Human genome-deriv
31	437.2	8.6	465	24	ABL81689	Human ovarian canc
32	429.6	8.4	819	19	AAV47512	Mouse neurotrophic
33	429.6	8.4	819	19	AAV22654	cDNA encoding muri
34	429.6	8.4	819	21	AAA39483	Murine NNT-1 cDNA.
35	429.6	8.4	819	24	ABK11649	Mouse cDNA encodin
36	429.4	8.4	432	24	ABL81658	Human ovarian canc
37	421.8	8.3	648	21	AAA88547	Mouse interieukin-
38	355.8	7.0	360	24	ABL79632	Human ovarian canc
39	330	6.5	342	20	AAK51546	Human secreted pro
40	329	6.5	340	21	AAI14792	Human secreted pro
41	324.6	6.4	396	20	AAI16162	Human cardiophi
42	250.6	4.9	269	24	ABL82404	Human ovarian canc
43	250.6	4.9	275	21	AAA42199	Human secreted exp
44	132.8	3.8	231	24	ABL79661	Human ovarian canc
45	165.2	3.2	283	22	AAH23145	Osteoarthritis tis

ALIGNMENTS

RESULT 1

AAA39482
ID AAA39482 standard; DNA; 5087 BP.

AC AAA39482;

XX 24-AUG-2000 (first entry)

DT Human NNT-1 DNA.

DE NNT-1; human; neurotrophic factor; neurotrophic; treatment;
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
KW anyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
KW retinopathy; immune disorder; hematopoietic disorder; ss.

OS Homo sapiens.

XX US6054294-A.

XX 25-APR-2000.

XX 12-DEC-1997; 97US-0988819.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M;

XX WPI; 2000-338492/29.


```
Db 3841 GGGGCTCATGGCTTCTGACTTCTGACCTTCTCCTTCTCGCTCCCTTCAAAACCTGCT 3900
Qy 3901 CCCATTGTGAGAGCCAGCCCTGTATGCAACACACTGTGTGAGCCAGGAGACAGAAGCTG 3960
Db 3901 CCCATTGTGAGAGCCAGCCCTGTATGCAACACACTGTGTGAGCCAGGAGACAGAAGCTG 3960
Qy 3961 TGAGCCTCTGGCCCTTCTGAGACCGGCTGGGGGTGTGATGGATCAGCCCTGTCTCTTC 4020
Db 3961 TGAGCCTCTGGCCCTTCTGAGACCGGCTGGGGGTGTGATGGATCAGCCCTGTCTCTTC 4020
Qy 4021 CCCACTCCCAAGGCTTACCGAGCTGGGAGGAGGTACAGTAGGCGCTGTCTGTCTCTG 4080
Db 4021 CCCACTCCCAAGGCTTACCGAGCTGGGAGGAGGTACAGTAGGCGCTGTCTGTCTCTG 4080
Qy 4081 TTTCTACAGGAAGTCATGCTCAGGAGGTGTGAAGTGGTTAGGTTGGTGCAGAGGGCT 4140
Db 4081 TTTCTACAGGAAGTCATGCTCAGGAGGTGTGAAGTGGTTAGGTTGGTGCAGAGGGCT 4140
Qy 4141 CATGGCTCTGCTTCTTGCTTACACTTGGCCAGTGGCCAGCCAGCCCTCAGGTGGCA 4200
Db 4141 CATGGCTCTGCTTCTTGCTTACACTTGGCCAGTGGCCAGCCAGCCCTCAGGTGGCA 4200
Qy 4201 CATCTGGAGGCGAGGGGTTGAGGGGCCACACACACATGCTTCTTGGGGTGAAGCCCT 4260
Db 4201 CATCTGGAGGCGAGGGGTTGAGGGGCCACACACACATGCTTCTTGGGGTGAAGCCCT 4260
Qy 4261 TTGGCTGCCCACTCTCTTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACA 4320
Db 4261 TTGGCTGCCCACTCTCTTGGATGGGTGTGCTCCCTTATCCCAATCACTCTATACA 4320
Qy 4321 TCCNATTCAGGAACAAACATGTGGCAATTTCTACACAAAAGAGATGAGATTAACAGTG 4380
Db 4321 TCCNATTCAGGAACAAACATGTGGCAATTTCTACACAAAAGAGATGAGATTAACAGTG 4380
Qy 4381 CAGGGTTGGGGTCTGCACTTGGAGTGGCTTATTAACAGAGAGAAATACTGAAAGCAC 4440
Db 4381 CAGGGTTGGGGTCTGCACTTGGAGTGGCTTATTAACAGAGAGAAATACTGAAAGCAC 4440
Qy 4441 AGGGGAGGAGACAGACAGACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500
Db 4441 AGGGGAGGAGACAGACAGACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500
Qy 4501 ACCGAGCTCAGCATCAGGACCTTGCTCGAATGTCTTCAGTATTAACGGTGCCTCTTC 4560
Db 4501 ACCGAGCTCAGCATCAGGACCTTGCTCGAATGTCTTCAGTATTAACGGTGCCTCTTC 4560
Qy 4561 TCTGCCCTTTCCAGGGTATCTGTGGGTGCGAGGTGGGAGGGCAACATAGCCAC 4620
Db 4561 TCTGCCCTTTCCAGGGTATCTGTGGGTGCGAGGTGGGAGGGCAACATAGCCAC 4620
Qy 4621 ACCACAGGATTTCTGAAAGTTTACATGAGTAGCATTTTGGGGGTGAGGGTGGCAGCT 4680
Db 4621 ACCACAGGATTTCTGAAAGTTTACATGAGTAGCATTTTGGGGGTGAGGGTGGCAGCT 4680
Qy 4681 CCCCAAGGCTTGGCCCGCCAGCCACCCATCATGCTTAAGTGTGTGTTATTAATAT 4740
Db 4681 CCCCAAGGCTTGGCCCGCCAGCCACCCATCATGCTTAAGTGTGTGTTATTAATAT 4740
Qy 4741 TTAATTAATGGAGATGTTATTAATAGATGATATTAATGAGATTTCTATTTCTGTA 4800
Db 4741 TTAATTAATGGAGATGTTATTAATAGATGATATTAATGAGATTTCTATTTCTGTA 4800
Qy 4801 TTAACAAATAAAGTCTTGGCCCGAGAACTTAGTCTTGTGCCAGCTTCCCTCTCTGG 4860
Db 4801 TTAACAAATAAAGTCTTGGCCCGAGAACTTAGTCTTGTGCCAGCTTCCCTCTCTGG 4860
Qy 4861 TGCTCATCAGACTCTTGCCACCCCTGGCTCCCACTTCCCTGTGCTCTGTGGAGTGC 4920
Db 4861 TGCTCATCAGACTCTTGCCACCCCTGGCTCCCACTTCCCTGTGCTCTGTGGAGTGC 4920
Qy 4921 ACAGAGCTCTGGAGAGGAGGCTCTTCTCTCCCGACATGGGGCGATGGCGCACCTCAGA 4980
Db 4921 ACAGAGCTCTGGAGAGGAGGCTCTTCTCTCCCGACATGGGGCGATGGCGCACCTCAGA 4980

Qy 4981 CTTATCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040
Db 4981 CTTATCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCCACACAGCTTC 5040
Qy 5041 TGTCACCCAGGTTTCCTTCAACCCACCTTGTGTAAGTCTTCCTCA 5087
Db 5041 TGTCACCCAGGTTTCCTTCAACCCACCTTGTGTAAGTCTTCCTCA 5087

RESULT 2
ABK11648
ID ABK11648 standard; DNA; 5087 BP.
XX
AC ABK11648;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human novel neurotrophic factor NNT1, genomic DNA.
XX
KW Human; ds; gene; NNT1; neurotrophic factor; IGE-related disease;
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
KW vascular stenosis; rheumatoid arthritis; psoriatic arthritis;
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
KW inflammatory bowel disease; transplant rejection; reproductive disorder;
KW graft versus host disease; infertility; miscarriage; preterm labour.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
misc_feature 137..138
/*tag= a
/ note= "Intervening, unsequenced region of more than 1KB"
XX
WO200215977-A2.
XX
28-FEB-2002.
XX
17-AUG-2001; 2001WO-US25906.
XX
18-AUG-2000; 2000US-226436P.
XX
16-AUG-2001; 2001US-0931704.
XX
(AMGE-) AMGEN INC.
XX
Senaldi G;
XX
WPI; 2002-280867/32.
XX
Treating Immunoglobulin E-related disease, modulating IgE levels in a
patient, preventing IGE-related disease and treating allergic diseases,
involves administering NNT-1 inhibitor to a patient
XX
Claim 2; Fig 2; 63pp; English.
XX
The invention relates to treating Immunoglobulin E (IGE)-related disease,
modulating IGE levels in a patient, preventing an IGE-related disease,
and treating allergic diseases, comprising administering a
therapeutically effective amount of novel neurotrophic factor (NNT)-1
inhibitor to a patient. Also included are a method of diagnosing an
IGE-related disease or susceptibility to an IGE-related disease, by
determining the presence or amount of expression of an NNT1 polypeptide
encoded by a NNT1 nucleotide sequence, its fragment or naturally
occurring variant, and diagnosing an IGE-related disease or
susceptibility of an IGE-related disease based on the presence or amount
of expression of the polypeptide and a pharmaceutical composition for use
in treating IGE-related disease, comprising the NNT1 inhibitor.
The NNT1 inhibitor is useful for preventing and treating IGE-related
disease, modulating IGE levels, and treating allergic diseases e.g.
Type I allergic disease, allergic rhinitis, eczema, dermatitis,
pollinosis, asthma, immune diseases and disorders, diseases involving
```


CC abnormal cell proliferation including cancer, arteriosclerosis and
CC immune system including rheumatoid arthritis, psoriatic arthritis,
CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,
CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
CC bowel disease, transplant rejection, and graft versus host disease, and
CC reproductive diseases and disorders including infertility, miscarriage,
CC preterm labour and delivery, and endometriosis. The present sequence
XX is the genomic DNA for human NNT1.

SQ Sequence 5087 BP; 992 A; 1746 C; 1191 G; 1158 T; 0 other;

Query Match 100.0%; Score 5087; DB 24; Length 5087;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AACCTGCCAGTGGGCTGGCGGATGGGATTATTAAAGCTTCCCGGAGCGCGGCTCGCC	60
Db	1	AACCTGGGAGTGGGCTGGCGGATGGGATTATTAAAGCTTCCCGGAGCGCGGCTCGCC	60
Qy	61	CTCCACCTCCGACGCTCCCGGAGGAGCGCGACCCGGCGGCGCCAGCCCGACCCCA	120
Db	61	CTCCACCTCCGACGCTCCCGGAGGAGCGCGACCCGGCGGCGCCAGCCCGACCCCA	120
Qy	121	TGGACCTCCGAGCAGGTTGAAACCCCAACTAGCCCTGCTCTTCAATACATGACAGCAG	180
Db	121	TGGACCTCCGAGCAGGTTGAAACCCCAACTAGCCCTGCTCTTCAATACATGACAGCAG	180
Qy	181	CGCCCATCTGATACCTAAACCGACCAAGTCACAGCCCTCAACCTCACCTCTGCTGCC	240
Db	181	CGCCCATCTGATACCTAAACCGACCAAGTCACAGCCCTCAACCTCACCTCTGCTGCC	240
Qy	241	CAGACCTCACCACCTCTTGTGGATCAAACTCAACCGCACTAAATCAACCAATCCCA	300
Db	241	CAGACCTCACCACCTCTTGTGGATCAAACTCAACCGCACTAAATCAACCAATCCCA	300
Qy	301	AGTCTAAACTAATCTGAACCTTTAAAGTAAACCCAGTCTTAAACCTAGCCCAAT	360
Db	301	AGTCTAAACTAATCTGAACCTTTAAAGTAAACCCAGTCTTAAACCTAGCCCAAT	360
Qy	361	GCCAAATATATCTACCTAGCCAAACCTTAACCTGCTTGGCAGTCCAAAGTGTCCACTG	420
Db	361	GCCAAATATATCTACCTAGCCAAACCTTAACCTGCTTGGCAGTCCAAAGTGTCCACTG	420
Qy	421	AATCTCACTTGTGCTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCCACAT	480
Db	421	AATCTCACTTGTGCTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCCACAT	480
Qy	481	CCCTCTTACAGCACCACCTGCGCTCTGACCTCTGATCTCTGGGATGTCCAAACT	540
Db	481	CCCTCTTACAGCACCACCTGCGCTCTGACCTCTGATCTCTGGGATGTCCAAACT	540
Qy	541	CTGAGTGCATCAGCCAAAGCCGACTCGTCAAAATGACCTCTCTCCCTTCTCTGTC	600
Db	541	CTGAGTGCATCAGCCAAAGCCGACTCGTCAAAATGACCTCTCTCCCTTCTCTGTC	600
Qy	601	CCACCTTGCAGGTGATGGAAGGCGCTCAITGAAGTCCAACTTTCCCACTTAACACC	660
Db	601	CCACCTTGCAGGTGATGGAAGGCGCTCAITGAAGTCCAACTTTCCCACTTAACACC	660
Qy	661	AAGAACGGGTGAACCTTCCACCTGCACTGCGTCCAGTCCCTGAGAGTGAGCACTAACTCTCT	720
Db	661	AAGAACGGGTGAACCTTCCACCTGCACTGCGTCCAGTCCCTGAGAGTGAGCACTAACTCTCT	720
Qy	721	CAATCTAACCCCACTACCTTCCACACTCAGGAATACATCTAGAAATATACCCAAA	780
Db	721	CAATCTAACCCCACTACCTTCCACACTCAGGAATACATCTAGAAATATACCCAAA	780
Qy	781	ACTAAGCCCAATAAGGAGCCGACCTTAAAGTGTCTAAACCTATACCTTGTCTTCTATGG	840
Db	781	ACTAAGCCCAATAAGGAGCCGACCTTAAAGTGTCTAAACCTATACCTTGTCTTCTATGG	840
Qy	841	GTGAGTCTGTCTTGGGGCGGCTCTCTCTGCTTCTCTCCCTTAGAGCTGAGTGTCT	900

Db	841	GTGAGTCTGTCTTGGGGCGGCTCTCTCTGCTTCTCTCCCTTAGAGCTGAGTGTCT	900
Qy	901	AGCCTGCCAGCTCTGACATGTCTGCTCCCACTCTGACTCCCTCAAGCTGCAAGTGG	960
Db	901	AGCCTGCCAGCTCTGACATGTCTGCTCCCACTCTGACTCCCTCAAGCTGCAAGTGG	960
Qy	961	GACTGGAAGACTGGGAGGAAGCTAGGGTAACTGGAACACAGGAGGTGCACTGACGT	1020
Db	961	GACTGGAAGACTGGGAGGAAGCTAGGGTAACTGGAACACAGGAGGTGCACTGACGT	1020
Qy	1021	CCCTAGGCTGGCCCGCTCCCTCCATGTACACATATATATATATATATATATATATAT	1080
Db	1021	CCCTAGGCTGGCCCGCTCCCTCCATGTATACACATATATATATATATATATATATAT	1080
Qy	1081	GCACATGTCGCAAGACTCTCTCAGCTGACACACAGATCCATTTCAAGTATCTACTGAT	1140
Db	1081	GCACATGTCGCAAGACTCTCTCAGCTGACACACAGATCCATTTCAAGTATCTACTGAT	1140
Qy	1141	AGACACTCATGCTGCCAAGTCTCTCATCTCTCAAAATATACATATATATATATATAT	1200
Db	1141	AGACACTCATGCTGCCAAGTCTCTCATCTCTCAAAATATACATATATATATATATAT	1200
Qy	1201	GTCTTGCAGGAGTGTTCCT	1260
Db	1201	GTCTTGCAGGAGTGTTCCT	1260
Qy	1261	TCACCCCGCCAGCCAGGCTGGGGAGACACACTGAGGGGCTGCCAGCTGCTTCCCT	1320
Db	1261	TCACCCCGCCAGCCAGGCTGGGGAGACACACTGAGGGGCTGCCAGCTGCTTCCCT	1320
Qy	1321	GTGTGGGCGCGGCGCTCATGCTCTCTGCTCCATCTGCTCCCTCTCTCTCTCTCTCTCT	1380
Db	1321	GTGTGGGCGCGGCGCTCATGCTCTCTGCTCCATCTGCTCCCTCTCTCTCTCTCTCTCT	1380
Qy	1381	GATGTTAGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCT	1440
Db	1381	GATGTTAGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCT	1440
Qy	1441	CACAGGGAGCCAGGCGCTGCTCCATCCAGAAACCTATGACCTCACCGCTACTCT	1500
Db	1441	CACAGGGAGCCAGGCGCTGCTCCATCCAGAAACCTATGACCTCACCGCTACTCT	1500
Qy	1501	GGAGCACCACTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1560
Db	1501	GGAGCACCACTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1560
Qy	1561	AGTTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGT	1620
Db	1561	AGTTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGT	1620
Qy	1621	TGATGGGAGGAGGAGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGT	1680
Db	1621	TGATGGGAGGAGGAGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGT	1680
Qy	1681	CCCTTCTCTCTTCT	1740
Db	1681	CCCTTCTCTCTTCT	1740
Qy	1741	GGCAGGCTCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1800
Db	1741	GGCAGGCTCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1800
Qy	1801	TCCCTCTGCGGCGCGGAGTCT	1860
Db	1801	TCCCTCTGCGGCGCGGAGTCT	1860
Qy	1861	GGCGGGGTGGGTGGGAGCAGAGGGGCGCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1920
Db	1861	GGCGGGGTGGGTGGGAGCAGAGGGGCGCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1920
Qy	1921	TGCCCCCAGACCTGGGGCGCTGCTCTGAGCCAGGGGCTCTCTCTCTCTCTCTCTCTCTCTCT	1980

CC inflammation. NNT-1 is also able to boost immunoreactivity and
CC antibody production following vaccination, and, since it inhibits
CC tumour necrosis factor production, it may also be useful for
CC treating sepsis. NNT-1 nucleic acid fragments are also used as
CC hybridisation probes in diagnostic assays. In addition, cells that
CC have been engineered to express NNT-1 can be implanted, or nucleic
CC acids are delivered in gene therapy vectors.

XX Sequence 5088 BP; 992 A; 1746 C; 1191 G; 1158 T; 1 other;

Query Match 99.8%; Score 5076; DB 19; Length 5088;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy 1 AACCTGCGAGTGGGCTGGCGATGGGATTAATTAAGCTTCGCGGAGCGCGGCTCGCC 60
Db 1 AACCTGCGAGTGGGCTGGCGATGGGATTAATTAAGCTTCGCGGAGCGCGGCTCGCC 60

Qy 61 CTCCTCCACTCCGCGAGCTCCGCGAGAGCGCGACCCCGCGCGCGCCAGCCCGCCCA 120
Db 61 CTCCTCCACTCCGCGAGCTCCGCGAGAGCGCGACCCCGCGCGCGCCAGCCCGCCCA 120

Qy 121 TGGACCTCCGAGCAGGTTGAACCCAAACTAGCCCTGCTCTTCAATAAGCAAGCAG 180
Db 121 TGGACCTCCGAGCAGGTTGAACCCAAACTAGCCCTGCTCTTCAATAAGCAAGCAG 180

Qy 181 CGCCCACTGATACCTTAACCGACCAAGTCACAGCCCTCCCACTCACCTCTGCCCTGCC 240
Db 181 CGCCCACTGATACCTTAACCGACCAAGTCACAGCCCTCCCACTCACCTCTGCCCTGCC 240

Qy 241 CAGACCTCACCATCCTTTG-TGGACTCAAACTCAACCGCACTAAATCAACCAATCCC 299
Db 241 CAGACCTCACCATCCTTTG-TGGACTCAAACTCAACCGCACTAAATCAACCAATCCC 300

Qy 300 AAGTCTAAACTAATCTGAACCTTTAAAGTAACCCAGTCTTTAAACTAACCTAGCCCAA 359
Db 301 AAGTCTAAACTAATCTGAACCTTTAAAGTAACCCAGTCTTTAAACTAACCTAGCCCAA 360

Qy 360 TGGCAATTAATCTACCTAGCAACCCCTTAAGTCTGCTTTGCGAGTCCCAAGTCTCCACT 419
Db 361 TGGCAATTAATCTACCTAGCAACCCCTTAAGTCTGCTTTGCGAGTCCCAAGTCTCCACT 420

Qy 420 GAATCTCACTTGGTCTCACTGAAATCCAGAAAGCAATATTTCCCACTGCCCAACA 479
Db 421 GAATCTCACTTGGTCTCACTGAAATCCAGAAAGCAATATTTCCCACTGCCCAACA 480

Qy 480 TCCCTCTTACAGCAACCAACCTGGCTCTGGACTCTGCTGATTCCTGGGATGTCGAAC 539
Db 481 TCCCTCTTACAGCAACCAACCTGGCTCTGGACTCTGCTGATTCCTGGGATGTCGAAC 540

Qy 540 TCTGGAGTGGCATCAGCCCAACCGCGACTCGTCAAACTCACTCTCCCTTCCTGTC 599
Db 541 TCTGGAGTGGCATCAGCCCAACCGCGACTCGTCAAACTCACTCTCCCTTCCTGTC 600

Qy 600 CCCACCTTGCAGGCTGATGGAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAACAC 659
Db 601 CCCACCTTGCAGGCTGATGGAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAACAC 660

Qy 660 CAAGAAAGGGGTGAACCTCCACACTGCCACCGTTCTCGAGAGTGAGCACTAAATCTCCT 719
Db 661 CAAGAAAGGGGTGAACCTCCACACTGCCACCGTTCTCGAGAGTGAGCACTAAATCTCCT 720

Qy 720 TCAATCTAACCCCACTCACTTCCACACTCAGGAATCACTCTAGATATACCCAA 779
Db 721 TCAATCTAACCCCACTCACTTCCACACTCAGGAATCACTCTAGATATACCCAA 780

Qy 780 AACTAAGCCCCATAAGCAGCCCGACCTTAGTGGTCTAAGCCCTATACCTTCTCTATG 839
Db 781 AACTAAGCCCCATAAGCAGCCCGACCTTAGTGGTCTAAGCCCTATACCTTCTCTATG 840

Qy 840 GGTGAGTCTGTTTGGGCGCGCTCTCTCTGCTCTGCTCTGAGACCCAGGGGCTCCCTTCCGCTC 899
Db 841 GGTGAGTCTGTTTGGGCGCGCTCTCTCTGCTCTGCTCTGAGACCCAGGGGCTCCCTTCCGCTC 900
```

```
Qy 900 CAGCCTGCCAGCTCTGACATGTGCTGTCTCCACCCTCTGACTCCCTCAAGCTGCAGTG 959
Db 901 CAGCCTGCCAGCTCTGACATGTGCTGTCTCCACCCTCTGACTCCCTCAAGCTGCAGTG 960

Qy 960 GGACTGGAAGACTGGCAGGAGCTAGGGTACAACCTGGAACACAGCAGGCTGCAGCTGAG 1019
Db 961 GGACTGGAAGACTGGCAGGAGCTAGGGTACAACCTGGAACACAGCAGGCTGCAGCTGAG 1020

Qy 1020 TCCTTAGGCTGGCCCGCTCCCTCCATGTATACACATATACATGTGGCACACACAGT 1079
Db 1021 TCCTTAGGCTGGCCCGCTCCCTCCATGTATACACATATACATGTGGCACACACAGT 1080

Qy 1080 GGCACACATGCCAAAGACTCTCTAGCTGACACACAGATCCATTTCAAGTATCTACTGA 1139
Db 1081 GGCACACATGCCAAAGACTCTCTAGCTGACACACAGATCCATTTCAAGTATCTACTGA 1140

Qy 1140 TAGACACTATCGGTGCAAGTCTCTCATCTCTCAACATACACATGCTCTCTTCTCTCC 1199
Db 1141 TAGACACTATCGGTGCAAGTCTCTCATCTCTCAACATACACATGCTCTCTTCTCTCC 1200

Qy 1200 CGTCTTGCCAGGAGTGTTCCTCCCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCCACC 1259
Db 1201 CGTCTTGCCAGGAGTGTTCCTCCCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCCACC 1260

Qy 1260 CTCACCCCCCAACCCAGCCCAAGGTGGGACAGACACCTGAGGGGCTGCCAGTGTCTCC 1319
Db 1261 CTCACCCCCCAACCCAGCCCAAGGTGGGACAGACACCTGAGGGGCTGCCAGTGTCTCC 1320

Qy 1320 CGTGTGGGCGGGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1379
Db 1321 CGTGTGGGCGGGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380

Qy 1380 GGATTTAGCGTGTCTGTGACCGTGTCTGCGACCTCTGCGACCTCTGCGAGTCTCAATC 1439
Db 1381 GGATTTAGCGTGTCTGTGACCGTGTCTGCGACCTCTGCGACCTCTGCGAGTCTCAATC 1440

Qy 1440 GCACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCGCTACC 1499
Db 1441 GCACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCGCTACC 1500

Qy 1500 TGGAGCAACAACTCGGAGCTTGGCTGGAGCTATGTCAGTATCCAGCTAGGAATCTGG 1559
Db 1501 TGGAGCAACAACTCGGAGCTTGGCTGGAGCTATGTCAGTATCCAGCTAGGAATCTGG 1560

Qy 1560 GAGTTGGGAGGAGTGAGGAGTTGGGGAAGACAGTCTCTAAACCGTGGAGGGTCTGGTAA 1619
Db 1561 GAGTTGGGAGGAGTGAGGAGTTGGGGAAGACAGTCTCTAAACCGTGGAGGGTCTGGTAA 1620

Qy 1620 ATGATGGGCTGAGGAGGGCTCTTTGGCTGCCACAGTCCCTCTGCTGGTCTATCTCT 1679
Db 1621 ATGATGGGCTGAGGAGGGCTCTTTGGCTGCCACAGTCCCTCTGCTGGTCTATCTCT 1680

Qy 1680 GCCCTTCCCTTATAGTGGCGCCCGCCCACTTCCCATCCCTGGCCCGCCAGGACTAGGATGT 1739
Db 1681 GCCCTTCCCTTATAGTGGCGCCCGCCCACTTCCCATCCCTGGCCCGCCAGGACTAGGATGT 1740

Qy 1740 GGGCAGGCTCGCACCCCGCTTGGCCCATTTGCCCACTGGCTGCCAGCCAGCCGCGCGC 1799
Db 1741 GGGCAGGCTCGCACCCCGCTTGGCCCATTTGCCCACTGGCTGCCAGCCAGCCGCGCGC 1800

Qy 1800 CTCCCTCTGGGCGCGGGAGTCTCTCTGTTTACACCGTGTGGTGTCTCTTGGCGC 1859
Db 1801 CTCCCTCTGGGCGCGGGAGTCTCTCTGTTTACACCGTGTGGTGTCTCTTGGCGC 1860

Qy 1860 GGGCGGGGTGGGTGGGACAGAGGGGCCCACTCCCATGCTGCTGCTCCAGTCCGCT 1919
Db 1861 GGGCGGGGTGGGTGGGACAGAGGGGCCCACTCCCATGCTGCTGCTCCAGTCCGCT 1920

Qy 1920 CTGCCCCCAGACCTGGGCGCTGCTGCTGCTGAGACCCAGGGGCTCCCTTCCGCTC 1979
Db 1921 CTGCCCCCAGACCTGGGCGCTGCTGCTGCTGAGACCCAGGGGCTCCCTTCCGCTC 1980
```


Db 4141 TCATGGCTCTGCTTCTTGGCTTACACACTTGGCCAGTGGCCACCCAGCCCTCAGGTGGC 4200
Qy 4200 ACATCTGGAGGCGAGGGTTGAGGGCCACCCACACATGCTTCTTGGGGTGAAGCCC 4259
Db 4201 ACATCTGGAGGCGAGGGTTGAGGGCCACCCACACATGCTTCTTGGGGTGAAGCCC 4260
Qy 4260 TTTGGTGGCCCACTCTCTTGGATGGGTGTGCTCCCTTATCCCAAACTCCTATATAC 4319
Db 4261 TTTGGTGGCCCACTCTCTTGGATGGGTGTGCTCCCTTATCCCAAACTCCTATATAC 4320
Qy 4320 ATCCAAATTCAGGAAACAAACATGGTGGCAATCTACACAAAGAGATGAGATTAACAGT 4379
Db 4321 ATCCAAATTCAGGAAACAAACATGGTGGCAATCTACACAAAGAGATGAGATTAACAGT 4380
Qy 4380 GCAGGGTTGGGTTCTCATTTGGAGTGGCCCTATAAACAGAGAAAATCTGAAGCA 4439
Db 4381 GCAGGGTTGGGTTCTCATTTGGAGTGGCCCTATAAACAGAGAAAATCTGAAGCA 4440
Qy 4440 CAGGGCAGGAGACAGACCCAGACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAA 4499
Db 4441 CAGGGCAGGAGACAGACCCAGACAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAA 4500
Qy 4500 AACCCGAGCTGAGCATCAGACCTTGGCTCGAATTTGCTTCCAGTATTACGGTGCCTCTT 4559
Db 4501 AACCCGAGCTGAGCATCAGACCTTGGCTCGAATTTGCTTCCAGTATTACGGTGCCTCTT 4560
Qy 4560 CTCTGCCCCCTTCCAGGGTATCTGTTGGTGGCAGCTGGGAGGACCACTAGCCA 4619
Db 4561 CTCTGCCCCCTTCCAGGGTATCTGTTGGTGGCAGCTGGGAGGACCACTAGCCA 4620
Qy 4620 CACCACAGGATTTCTCAAAATTTACATGCAATGAGTAGATTTGGGGTGTAGGGTGGCAGC 4679
Db 4621 CACCACAGGATTTCTCAAAATTTACATGCAATGAGTAGATTTGGGGTGTAGGGTGGCAGC 4680
Qy 4680 TCCCCAAGGCCCTGCCCCCAGCCCCACCACTCATGCTAAAGTGTGTGTAATA 4739
Db 4681 TCCCCAAGGCCCTGCCCCCAGCCCCACCACTCATGCTAAAGTGTGTGTAATA 4740
Qy 4740 TTTATTATTGGAGATTTATTATTATTAGATGATATTATTGCAGATTTCTATTCTTGT 4799
Db 4741 TTTATTATTGGAGATTTATTATTATTAGATGATATTATTGCAGATTTCTATTCTTGT 4800
Qy 4800 ATTAACAAATAAATGCTTGGCCAGAACCTTAGTCTCTTTGCCAGCCTCACCCCTCTG 4859
Db 4801 ATTAACAAATAAATGCTTGGCCAGAACCTTAGTCTCTTTGCCAGCCTCACCCCTCTG 4860
Qy 4860 GTGCTCATAGACTCTTGGCCACCCCTGGCTCCACTCCCTGCTTCCCTCTGGTGGAGCTG 4919
Db 4861 GTGCTCATAGACTCTTGGCCACCCCTGGCTCCACTCCCTGCTTCCCTCTGGTGGAGCTG 4920
Qy 4920 CACAGAGCTCTGGGAAGAGGCCCTCTTCTCCCGCACTGGGGCGATGGGGCGACCTCAG 4979
Db 4921 CACAGAGCTCTGGGAAGAGGCCCTCTTCTCCCGCACTGGGGCGATGGGGCGACCTCAG 4980
Qy 4980 ACTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCCAACAGCTT 5039
Db 4981 ACTTACCCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCTCTCCCAACAGCTT 5040
Qy 5040 CTGTCCACCCAGGTTTCCCTCACCCGACCTTGTGTAAGTCTTCTCA 5087
Db 5041 CTGTCCACCCAGGTTTCCCTCACCCGACCTTGTGTAAGTCTTCTCA 5088

RESULT 4

AAV22653

ID AAV22653 standard; DNA; 5088 BP.

XX AC

AAV22653;

XX AC

13-JUL-1998 (first entry)

XX DE

Human genomic DNA encoding neurotrophic factor NNT-1.

XX

Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
treatment; neurological disease; degeneration; Parkinson's disease;
amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.
Homo sapiens.
Key Location/Qualifiers
misc_feature 138
/*tag= a
/note= "represents intervening unsequenced region
of 1 Kb"

U85741772-A.

21-APR-1998.

03-FEB-1997; 97US-0792019.

03-FEB-1997; 97US-0792019.

(AMGE-) AMGEN INC.

Chang M;

WPI; 1998-260526/23.

Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
useful for stimulating growth of motor and sympathetic neurons

Disclosure; Fig 2; 41pp; English.

The present sequence encodes a human neurotrophic factor, designated
NNT-1, which is capable of stimulating growth of motor or sympathetic
neurons. The NNT-1 protein is useful in the treatment of neurological
diseases characterized by the degeneration and death of particular
classes of neurons. These diseases specifically include Parkinson's
disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
stroke and various degenerative disorders affecting vision.

Sequence 5088 BP; 992 A; 1746 C; 1191 G; 1158 T; 1 other;

Query Match 99.8%; Score 5076; DB 19; Length 5088;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 AACCTGCGAGTGGGCTGGCGGATGGGATTTAAAGTCTTCCGAGCGCGGCTCGCC 60

Db 1 AACCTGCGAGTGGGCTGGCGGATGGGATTTAAAGTCTTCCGAGCGCGGCTCGCC 60

Qy 61 CTCCCATCTCCGAGCTCGGGAGAGGAGCGCACCGCGCGCCAGCCAGCCCA 120

Db 61 CTCCCATCTCCGAGCTCGGGAGAGGAGCGCACCGCGCGCCAGCCAGCCCA 120

Qy 121 TGGACCTCCGAGCAGT-TGMAAACCCAACTAGCCTCTTCTTATAACATCAAGCA 179

Db 121 TGGACCTCCGAGCAGTNTGAAACCCAACTAGCCTCTTCTTATAACATCAAGCA 180

Qy 180 GGGCCCATCTGTATACCTAAACGACCAAGTCACAGCCCTCCAACTACCTCTGCTGC 239

Db 181 GGGCCCATCTGTATACCTAAACGACCAAGTCACAGCCCTCCAACTACCTCTGCTGC 240

Qy 240 CAGACCTCACCACATCTTGTGACCTCAACCTCAACCTCAACCTCAACCTCAACCT 299

Db 241 CAGACCTCACCACATCTTGTGACCTCAACCTCAACCTCAACCTCAACCTCAACCT 300

Qy 300 AAGTCTAAACTAATCTGAAACTTTTAAAGTAACTTAAAGTAACTTAAAGTAACT 359

Db 301 AAGTCTAAACTAATCTGAAACTTTTAAAGTAACTTAAAGTAACTTAAAGTAACT 360

Qy 360 TGCCAAATTAATCTACCCCTAGCCAAACCTTAACTGCTTCCAGTCCAAAGTGTCCACT 419

Db 361 TGCCAAATTAATCTACCCCTAGCCAAACCTTAACTGCTTCCAGTCCAAAGTGTCCACT 420

QY 420 GAATCCTCAGCTTGGTCTCTCACTGAAATCCAGAAAGCAATATTTCCCACTGCCCCA 479
Db 421 GAATCCTCAGCTTGGTCTCTCACTGAAATCCAGAAAGCAATATTTCCCACTGCCCCA 480
QY 480 TCCCTCTTTACAGCAACCAACCTGCTCTGACCTCTGCTGATCTCTGGATGTCACAA 539
Db 481 TCCCTCTTTACAGCAACCAACCTGCTCTGACCTCTGCTGATCTCTGGATGTCACAA 540
QY 540 TCTGAGTGCATACAGCAACCAACCTGCTCTGAAATGCACTCTCTCCCTTCTCTGTC 599
Db 541 TCTGAGTGCATACAGCAACCAACCTGCTCTGAAATGCACTCTCTCCCTTCTCTGTC 600
QY 600 CCCACCTTGCAGCTGATGGAAGAGCCCTCATTTGAAGTCAACTTTTCCCACTTAACAC 659
Db 601 CCCACCTTGCAGCTGATGGAAGAGCCCTCATTTGAAGTCAACTTTTCCCACTTAACAC 660
QY 660 CAAGAACGGGTGAACCTCCACACTGCCACCGTTCCCTGAGAGTGAGCACTAAATCTCT 719
Db 661 CAAGAACGGGTGAACCTCCACACTGCCACCGTTCCCTGAGAGTGAGCACTAAATCTCT 720
QY 720 TCAATCTAACCCCAACCTTACATTTCCCACTCAGGAATCAATCTAGAAATATACCCAA 779
Db 721 TCAATCTAACCCCAACCTTACATTTCCCACTCAGGAATCAATCTAGAAATATACCCAA 780
QY 780 AACTAAGCCCCATAGGAGCCCGACCTAGTGTCTTCCCTTAGAGCTGACCTGTGCT 839
Db 781 AACTAAGCCCCATAGGAGCCCGACCTAGTGTCTTCCCTTAGAGCTGACCTGTGCT 840
QY 840 GGTGAGTCTGTCTTTGGGGCCCGCTCTCTCTGCTCTCTCCCTTAGAGCTGACCTGTGCT 899
Db 841 GGTGAGTCTGTCTTTGGGGCCCGCTCTCTCTGCTCTCTCCCTTAGAGCTGACCTGTGCT 900
QY 900 CAGCTTGCAGCTCTGACATGTGTCTTCCCAACCTCTGACTCCCTCAAGCTGCAGTG 959
Db 901 CAGCTTGCAGCTCTGACATGTGTCTTCCCAACCTCTGACTCCCTCAAGCTGCAGTG 960
QY 960 GGACTGGAAGACTGCGAGAGCTAGGTACAACTGGAAACAGAGCTGACCTGAC 1019
Db 961 GGACTGGAAGACTGCGAGAGCTAGGTACAACTGGAAACAGAGCTGACCTGAC 1020
QY 1020 TCCCTAGGCTGGCCCCCTCTCCATGTACACATATATATGTTGGCACACACACAGT 1079
Db 1021 TCCCTAGGCTGGCCCCCTCTCCATGTACACATATATATGTTGGCACACACACAGT 1080
QY 1080 GGCAACATGCGCAAGACTCTCTGAGTGAACACAGATCCATTTCTCAAGTATCTACTGA 1139
Db 1081 GGCAACATGCGCAAGACTCTCTGAGTGAACACAGATCCATTTCTCAAGTATCTACTGA 1140
QY 1140 TAGACACTCATGCGTGCAGGCTCTCTCACTCCCAACATACATGCTCTCTCTCTCTCC 1199
Db 1141 TAGACACTCATGCGTGCAGGCTCTCTCACTCCCAACATACATGCTCTCTCTCTCTCC 1200
QY 1200 CGTCTTGCAGAGTGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1259
Db 1201 CGTCTTGCAGAGTGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1260
QY 1260 CTCACCCCAACAGCCAGCTGAGTGGGAACACCTGAGGGGTGCGAGTGTCTTCC 1319
Db 1261 CTCACCCCAACAGCCAGCTGAGTGGGAACACCTGAGGGGTGCGAGTGTCTTCC 1320
QY 1320 CGTGTGGGCGCGGCGCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1379
Db 1321 CGTGTGGGCGCGGCGCTCATGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
QY 1380 GGATGTTAGCTGTGTCAGTGTCTGGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1439
Db 1381 GGATGTTAGCTGTGTCAGTGTCTGGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
QY 1440 GCAGGGGACCCAGGGCTGGCCCTTCCATCCAGAAACCTATGACCTGACCCGCTTACC 1499
Db 1441 GCAGGGGACCCAGGGCTGGCCCTTCCATCCAGAAACCTATGACCTGACCCGCTTACC 1500
QY 1500 TGGAGCAACCACTCGAGCTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGG 1559

Db 1501 TGAGCACCAACTCCGACCTTGGCTGGGACCTATGTGAGTATCCAGCGTAGGAATCTGG 1560
QY 1560 GAGTTGGGAGGAGTGGAGTGGGAAAGACAGTCTTAAACCGTGGAGGGTCTGGTAA 1619
Db 1561 GAGTTGGGAGGAGTGGAGTGGGAAAGACAGTCTTAAACCGTGGAGGGTCTGGTAA 1620
QY 1620 ATGATGGGAGGAGGGGCTCTTTGGCTCCCAACAGTCCCGCTGCTCTCTCTCTCTCTCT 1679
Db 1621 ATGATGGGAGGAGGGGCTCTTTGGCTCCCAACAGTCCCGCTGCTCTCTCTCTCTCTCT 1680
QY 1680 GGCCTTCCCTTTAGTGGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1739
Db 1681 GGCCTTCCCTTTAGTGGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
QY 1740 GGGCAGGCTCCACCCCGCTTGGCCATTTGGCCATTTGGCTCCCAACAGTCCCGCTGCTCTCTCT 1799
Db 1741 GGGCAGGCTCCACCCCGCTTGGCCATTTGGCCATTTGGCTCCCAACAGTCCCGCTGCTCTCTCT 1800
QY 1800 CTCCCTCTGGGGCCGGGAGTCT 1859
Db 1801 CTCCCTCTGGGGCCGGGAGTCT 1860
QY 1860 GGGCGGGTGGGTGGGACAGAGGGGCCCACTTCCATGCTGCTGCTCTCTCTCTCTCTCTCTCT 1919
Db 1861 GGGCGGGTGGGTGGGACAGAGGGGCCCACTTCCATGCTGCTGCTCTCTCTCTCTCTCTCTCT 1920
QY 1920 CTGCCCCCAGACTCGGGCCCTGCTCTGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCTCT 1979
Db 1921 CTGCCCCCAGACTCGGGCCCTGCTCTGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCTCT 1980
QY 1980 TCCCATCTAGCTGGGCTCTCTAGGGGGTCTAGGGGAGAGGGGCTGTAGGGAAACCCAG 2039
Db 1981 TCCCATCTAGCTGGGCTCTCTAGGGGGTCTAGGGGAGAGGGGCTGTAGGGAAACCCAG 2040
QY 2040 GCAGTAGTGGAGGGGTTTAGGGTGGATGGAGTTATGCTGTAAGGATTTGGGGGTG 2099
Db 2041 GCAGTAGTGGAGGGGTTTAGGGTGGATGGAGTTATGCTGTAAGGATTTGGGGGTG 2100
QY 2100 GTCCAGAGTGTTCAGAGAGCCAGGAGAGAGGAGAGGGTTCGAGAGGCCGAGGAC 2159
Db 2101 GTCCAGAGTGTTCAGAGAGCCAGGAGAGAGGAGAGGGTTCGAGAGGCCGAGGAC 2160
QY 2160 CATGGGAAACCCGCT 2219
Db 2161 CATGGGAAACCCGCT 2220
QY 2220 GCCAGGAAAGAAAGGAGAGAGTGGCGGGGAGCTGGCTCCAGCCCCAGGATACAC 2279
Db 2221 GCCAGGAAAGAAAGGAGAGAGTGGCGGGGAGCTGGCTCCAGCCCCAGGATACAC 2280
QY 2280 GAGGAAATTTAGTGTCT 2339
Db 2281 GAGGAAATTTAGTGTCT 2340
QY 2340 ATCCAGGCT 2399
Db 2341 ATCCAGGCT 2400
QY 2400 TTCCCTGGGCCCCAGCGCT 2459
Db 2401 TTCCCTGGGCCCCAGCGCT 2460
QY 2460 CATGCTGTCTTCCATAGCT 2519
Db 2461 CATGCTGTCTTCCATAGCT 2520
QY 2520 TCTGCAACCTGTCTCTATCAGCTGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2579
Db 2521 TCTGCAACCTGTCTCTATCAGCTGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2580
QY 2580 CTCCCAAGCCCTCAGCTGTGGGCT 2639

QY 4203 TCTGAGGCGAGGGTTGAGGGGCCACACACATGCTTTCTGGGTGAAGCCCTTT 4262
Db 1153 TCTGAGGCGAGGGTTGAGGGGCCACACACATGCTTTCTGGGTGAAGCCCTTT 1212
QY 4263 GGCTCCGCCACTCTCTTGGATGGGTGTTGCTCCCTTATCCCAAAATCACTCTATACATC 4322
Db 1213 GGCTCCGCCACTCTCTTGGATGGGTGTTGCTCCCTTATCCCAAAATCACTCTATACATC 1272
QY 4323 CAATTCAGGAACAAACATGTTGGCAATTTCTACAAAAAGAGATGAGATTAACAGTGCA 4382
Db 1273 CAATTCAGGAACAAACATGTTGGCAATTTCTACAAAAAGAGATGAGATTAACAGTGCA 1332
QY 4383 GGGTGGGGTCTGCAATGAGGTGCCCTATAAACAGAGAGAAAAATCTGAAAGCACAG 4442
Db 1333 GGGTGGGGTCTGCAATGAGGTGCCCTATAAACAGAGAGAAAAATCTGAAAGCACAG 1392
QY 4443 GGGCAGGACAGACACAGACCCAGAGCTCTCCAAAGCACAGAGTGCGCAAAACAAAC 4502
Db 1393 GGGCAGGACAGACACAGACCCAGAGCTCTCCAAAGCACAGAGTGCGCAAAACAAAC 1452
QY 4503 CCGAGCTGAGCATCAGGACCTTGCCTCGAATTTCTTCCAGTATTACGGTGCCTCTTCTC 4562
Db 1453 CCGAGCTGAGCATCAGGACCTTGCCTCGAATTTCTTCCAGTATTACGGTGCCTCTTCTC 1512
QY 4563 TGCCCCCTTTCCAGGGTATCTGTGGGTTGCCAGGCTGGGGAGGCAACCATAGCCACAC 4622
Db 1513 TGCCCCCTTTCCAGGGTATCTGTGGGTTGCCAGGCTGGGGAGGCAACCATAGCCACAC 1572
QY 4623 CACAGGATTTCTGAAAGTTTACAATGSCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCC 4682
Db 1573 CACAGGATTTCTGAAAGTTTACAATGSCAGTAGCATTTTGGGGTGTAGGGTGGCAGCTCC 1632
QY 4683 CCAAGCCCTGCCCCCAGCCACCCACCTCATGACTCTAAGTGTGTTGTTAATATTT 4742
Db 1633 CCAAGCCCTGCCCCCAGCCACCCACCTCATGACTCTAAGTGTGTTGTTAATATTT 1692
QY 4743 ATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTGAGATTTCTATTCTTGATT 4802
Db 1693 ATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTGAGATTTCTATTCTTGATT 1752
QY 4803 AACAAATAAAATGCTTGGCCCCAGAAC 4828
Db 1753 AACAAATAAAATGCTTGGCCCCAGAAC 1778

RESULT 6

AX16161
ID AX16161 standard; DNA; 1710 BP.

XX AC AX16161;

XX AC AX16161;

DT 28-APR-1999 (first entry)

DE Human cardiotrophin-like cytokine encoding DNA.

XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;
KW CLC; IL-6; diagnosis; detection; immune system-related disorder;
KW cancer; cardiac disorder; heart failure; hypertension; cancer;
KW autoimmune disorder; infection; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FH CDS 46..723
FT /*tag= a

FT sig_peptide 46..129

FT /*tag= b

FT mat_peptide 130..720

FT /*tag= c

FT /product= "cardiotrophin-like cytokine"

XX PN WO9900415-A1.

XX 07-JAN-1999.
XX 29-JUN-1998; 98WO-US13129.
XX 30-JUN-1997; 97US-0051311.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Shi Y;
XX WPI; 1999-095678/38.
XX P-PSDB; AAW94466.
XX New isolated cardiotrophin-like cytokine nucleic acid - used to
XX develop products for treating cardiac and immune system disorders,
XX e.g. heart failure, hypertension, cancers, autoimmune disorders and
XX PT infections
XX Claim 2; Fig 1; 103pp; English.
XX The present invention relates to a novel cardiotrophin-like cytokine
XX (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine
XX family. The present sequence encodes the human CLC protein. The
XX present invention also describes screening methods for identifying
XX agonists and antagonists of CLC activity, as well as methods for
XX detecting cardiac and immune system-related disorders and
XX therapeutic methods for treating cardiac and immune system-related
XX disorders, e.g. heart failure, hypertension, cancers, autoimmune
XX disorders and infections.
XX SQ Sequence 1710 BP; 370 A; 530 C; 448 G; 362 T; 0 other;

Query Match 28.8%; Score 1463.4; DB 20; Length 1710;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACTCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGG 3422
Db 227 ATCTGAACCTACTCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGG 286
QY 3423 CAGAGACTTGTCCCGAGGCCACTGTGTGACTTTGGAGGTGTGGGAGGCTCAATGACAAAC 3482
Db 287 CAGAGACTTGTCCCGAGGCCACTGTGTGACTTTGGAGGTGTGGGAGGCTCAATGACAAAC 346
QY 3483 TGCGGCTGACCCAGAACCTACGAGGCTACAGCCACTTCTGTGTTACTTGTGGCTCA 3542
Db 347 TGCGGCTGACCCAGAACCTACGAGGCTACAGCCACTTCTGTGTTACTTGTGGCTCA 406
QY 3543 ACCGTGAGGCTGCCACTGCTGAGCTGCGCGCAGCTTGGCCCACTTCTGACCAGCTCC 3602
Db 407 ACCGTGAGGCTGCCACTGCTGAGCTGCGCGCAGCTTGGCCCACTTCTGACCAGCTCC 466
QY 3603 AGGCGCTGCTGGGCGAGCATTTGGGCGTCTATGGAGCTTGGGCTACCCACTGCCCCAGC 3662
Db 467 AGGCGCTGCTGGGCGAGCATTTGGGCGTCTATGGAGCTTGGGCTACCCACTGCCCCAGC 526
QY 3663 CGCTGCTGGGAGCTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTACTTCTCTCAGA 3722
Db 527 CGCTGCTGGGAGCTGAACCCACTTGGACTCTCTGGCCCTGCCACAGTACTTCTCTCAGA 586
QY 3723 AGATGGAGCACTTCTGGCTGCTGAAGGAGCTGCAGACCTTGGCTGTGGCGCTCGGCCAAGG 3782
Db 587 AGATGGAGCACTTCTGGCTGCTGAAGGAGCTGCAGACCTTGGCTGTGGCGCTCGGCCAAGG 646
QY 3783 ACTTCAACCGGCTCAAGAAAGATGAGGCTTCCAGAGCTGCAGTCACTGACCTGG 3842
Db 647 ACTTCAACCGGCTCAAGAAAGATGAGGCTTCCAGAGCTGCAGTCACTGACCTGG 706
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3902
Db 707 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 766

CC codes for a novel neurotrophic factor, designated NNT-1 (see
CC AAV29715), that is a growth factor for neurons and for B or T cells.
CC It was obtained from a T-cell lymphoma cDNA library by expressed
CC sequence tag analysis on the basis of homology to CNTF. The
CC isolated NNT-1 cDNA was used as probe to isolated NNT-1 genomic
CC DNA (see AAV47511). Vectors containing the cDNA or genomic DNA and
CC host cells are provided for use in the production of NNT-1
CC polypeptides. These are used to treat: (i) neurological or
CC immunological diseases, specifically Alzheimer's, Parkinson's or
CC Huntington's diseases, amyotrophic lateral sclerosis,
CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
CC degeneration of the neural retina, or conditions characterised by T
CC or B cell defects, e.g. common variable immunodeficiency (CVID),
CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
CC inflammation. NNT-1 is also able to boost immunoreactivity and
CC antibody production following vaccination, and, since it inhibits
CC tumour necrosis factor production, it may also be useful for
CC treating sepsis. NNT-1 nucleic acid fragments are also used as
CC hybridisation probes in diagnostic assays. In addition, cells that
CC have been engineered to express NNT-1 can be implanted, or nucleic
CC acids are delivered in gene therapy vectors.
XX
SQ Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;
Query Match 10.3%; Score 523.4; DB 19; Length 797;
Best Local Similarity 99.8%; Pred. No. 5.1e-120;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330
QY 3423 CAGAGACTCTGCCAGGGGCGACTGTGACTTGAGGTGGCGAAGCCTCAATGACAAAC 3482
Db 331 CAGAGACTCTGCCAGGGGCGACTGTGACTTGAGGTGGCGAAGCCTCAATGACAAAC 390
QY 3483 TCGGCTGACCCAGAACTACGAGGCTTGTGACTTGAGGTGGCGAAGCCTCAATGACAAAC 3542
Db 391 TCGGCTGACCCAGAACTACGAGGCTTGTGACTTGAGGTGGCGAAGCCTCAATGACAAAC 450
QY 3543 ACCGTGAGGCTGACCTGTGACTGTGGCGAGCCTGGCCACTTTGTGACCAAGCCTCC 3602
Db 451 ACCGTGAGGCTGACCTGTGACTGTGGCGAGCCTGGCCACTTTGTGACCAAGCCTCC 510
QY 3603 AGGGCTGTGGGACACTTGGGCGGTATGGAGCTCTGGGCTACCCACTGCCCGAGC 3662
Db 511 AGGGCTGTGGGACACTTGGGCGGTATGGAGCTCTGGGCTACCCACTGCCCGAGC 570
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTCTTGGCCCTGCCAGTGACTTCTCTCCAGA 3722
Db 571 CGCTGCTGGGACTGAACCCACTTGGACTCTTGGCCCTGCCAGTGACTTCTCTCCAGA 630
QY 3723 AGATGAGACACTTCTGGCTGCTGAAGAGAGTGCAGACCTGGCTGTGGCGTGGCCCAAGG 3782
Db 631 AGATGAGACACTTCTGGCTGCTGAAGAGAGTGCAGACCTGGCTGTGGCGTGGCCCAAGG 690
QY 3783 ACTTCAACCGGCTCAAGAGAAAGATGAGAGCTCCAGAGCTGAGTCAACCTGACCTGG 3842
Db 691 ACTTCAACCGGCTCAAGAGAAAGATGAGAGCTCCAGAGCTGAGTCAACCTGACCTGG 750
QY 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCT 3887
Db 751 GGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCTGCTTCTGCT 795
RESULT 11
AAV22652
ID AAV22652 standard; cDNA; 797 BP.
XX
AC AAV22652;
XX
DT 13-JUL-1998 (first entry)

XX cDNA encoding human neurotrophic factor NNT-1.
DE Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
XX treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.
KW
XX Homo sapiens.
XX
XX Key Location/Qualifiers
PH CDS 90..767
FT /*tag= a
FT sig_peptide 90..170
FT /*tag= b
FT mat_peptide 171..764
FT /*tag= c
XX
XX US5741772-A.
PN 21-APR-1998.
XX
XX 03-FEB-1997; 97US-0792019.
PF
XX 03-FEB-1997; 97US-0792019.
PR
XX (AMGE-) AMGEN INC.
XX
XX Chang M;
PI
XX WPI; 1998-260526/23.
DR P-PSDB; AAW56141.
XX
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
PT useful for stimulating growth of motor and sympathetic neurons
XX
XX Disclosure; Fig 1; 41pp; English.
PS
XX The present sequence encodes a human neurotrophic factor, designated
CC NNT-1, which is capable of stimulating growth of motor or sympathetic
CC neurons. The NNT-1 protein is useful in the treatment of neurological
CC diseases characterised by the degeneration and death of particular
CC classes of neurons. These diseases specifically include Parkinson's
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
CC stroke and various degenerative disorders affecting vision.
XX
SQ Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;
Query Match 10.3%; Score 523.4; DB 19; Length 797;
Best Local Similarity 99.8%; Pred. No. 5.1e-120;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 330
QY 3423 CAGAGACTCTGCCAGGGGCGACTGTGACTTGAGGTGGCGAAGCCTCAATGACAAAC 3482
Db 331 CAGAGACTCTGCCAGGGGCGACTGTGACTTGAGGTGGCGAAGCCTCAATGACAAAC 390
QY 3483 TCGGCTGACCCAGAACTACGAGGCTTGTGACTTGAGGTGGCGAAGCCTCAATGACCTCA 3542
Db 391 TCGGCTGACCCAGAACTACGAGGCTTGTGACTTGAGGTGGCGAAGCCTCAATGACCTCA 450
QY 3543 ACCGTGAGGCTGACCTGTGACTGTGGCGAGCCTGGCCACTTCTCTCCAGAGCCTCC 3602
Db 451 ACCGTGAGGCTGACCTGTGACTGTGGCGAGCCTGGCCACTTCTCTCCAGAGCCTCC 510
QY 3603 AGGGCTGTGGGACACTTGGGCGGTATGGAGCTCTGGGCTACCCACTGCCCGAGC 3662
Db 511 AGGGCTGTGGGACACTTGGGCGGTATGGAGCTCTGGGCTACCCACTGCCCGAGC 570
QY 3663 CGCTGCTGGGACTGAACCCACTTGGACTCTTGGCCCTGCCAGTGACTTCTCTCCAGA 3722

CDS 90..767
/*tag= a
/product= "NNT1"

WO200215977-A2.
28-FEB-2002.
17-AUG-2001; 2001WO-US25906.
18-AUG-2000; 2000US-226436P.
16-AUG-2001; 2001US-0931704.
(ANGE-) ANGEN INC.
Senaldi G;
WPI; 2002-280867/32.
P-PSDB; AAU78176.

Treating Immunoglobulin E-related disease, modulating IgE levels in a patient, preventing IgE-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient -
Claim 2; Fig 1; 63pp; English.

The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor.

The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular restenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence encodes human NNT1.

Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;
Query Match 10.3%; Score 523.4; DB 24; Length 797;
Best Local Similarity 99.8%; Pred. No. 5,1e-120;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 3422
Db 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 330

QY 3423 CAGAGACTGTGCCAGGGCCACCTGTGACCTTGGAGGTGCGGAGCCTCAATGACAAAC 3482
Db 331 CAGAGACTGTGCCAGGGCCACCTGTGACCTTGGAGGTGCGGAGCCTCAATGACAAAC 390

QY 3483 TGGCGGTGACCCAGAACTACGAGGCCCTTACAGCCACCTTCTGTGTACTTGGGTGGCCCTCA 3542
Db 391 TGGCGGTGACCCAGAACTACGAGGCCCTTACAGCCACCTTCTGTGTACTTGGGTGGCCCTCA 450

QY 3543 ACCGTGAGGTGACCTGTGAGTGGCGCGAGCCTGGCCACTTCTGACACGAGCCTCC 3602
Db 451 ACCGTGAGGTGACCTGTGAGTGGCGCGAGCCTGGCCACTTCTGACACGAGCCTCC 510

QY 3603 AGGCGCTGTGGGAGCAGCATTTGCGGGCGTCTATGGCAGCTCTGGGCTACCCACTGCCCCAGC 3662
Db 511 AGGCGCTGTGGGAGCAGCATTTGCGGGCGTCTATGGCAGCTCTGGGCTACCCACTGCCCCAGC 570

QY 3663 CGTGGCTGGGAGCTGAACCCACATTTGGACTCTTGGCCCTGCCAGTCACTTCTCCAGA 3722
Db 571 CGTGGCTGGGAGCTGAACCCACATTTGGACTCTTGGCCCTGCCAGTCACTTCTCCAGA 630

QY 3723 AGATGGACGACTTCTGGCTGCTGAAGAGCTGCAGAGCTGCTGTGGCGCTCGGCAAGG 3782
Db 631 AGATGGACGACTTCTGGCTGCTGAAGAGCTGCAGAGCTGCTGTGGCGCTCGGCAAGG 690

QY 3783 ACTTCAACCGGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTGCAGTCACTTCCGACCTGG 3842
Db 691 ACTTCAACCGGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTGCAGTCACTTCCGACCTGG 750

QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTGGCTCCGCC 3887
Db 751 GGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTGGCTCCGCC 795

RESULT 14

ABA09140
ID ABA09140 standard; cDNA; 968 BP.

XX AC ABA09140;

XX DT 11-JAN-2002 (first entry)

XX DE Human cardiostrophin-like cytokine homologue cDNA, SEQ ID NO:916.

XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer; ss.

XX OS Homo sapiens.

XX PN WO200157188-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US03800.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX PI WPI; 2001-457740/49.

XX PR P-PSDB; ABB11896.

XX DR Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX PS Claim 1; Page 793-794; 1963pp; English.

XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a

nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoietic regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.

SQ Sequence 968 BP; 179 A; 353 C; 244 G; 192 T; 0 other;

Query Match	9.7%; Score 495.4; DB 22; Length 968;
Best Local Similarity	99.8%; Pred. No. 5.4e-113;
Matches 496; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY 3363	AGCTGMACTACTCGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGGG 3422
Db	
QY 472	ATCTGAACACTACTCGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGGGG 531
Db	
QY 3423	CAGAGACTCTGCCAGGGCCACTGTGTGACTTTGGAGGTCTGCGGAAGCCTCAATGACAAAC 3482
Db	
QY 532	CAGAGACTCTGCCAGGGCCACTGTGTGACTTTGGAGGTGCGGAAGCCTCAATGACAAAC 591
Db	
QY 3483	TGCGGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTACTTGGCTGGCCCTCA 3542
Db	
QY 592	TGCGGCTGACCCAGAACTACGAGGCCCTACAGCCACCTTCTGTGTACTTGGCTGGCCCTCA 651
Db	
QY 3543	ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCTTGGGCCCACTTCTGCAACGAGCTCC 3602
Db	
QY 652	ACCGTCAGGCTGCCACTGCTGAGCTGCGCCGAGCTTGGGCCCACTTCTGCAACGAGCTCC 711
Db	
QY 3603	AGGCGCTGCTGGGCGAGCAATTCGGGGCGTCATGGCAGCTCTGGGCTACCCACTGCCCGAC 3662
Db	
QY 712	AGGCGCTGCTGGGCGAGCAATTCGGGGCGTCATGGCAGCTCTGGGCTACCCACTGCCCGAC 771
Db	
QY 3663	CGCTGCCTGGGACTGAAACCCACTTTGGACTCTCGGCCCTGCCACAGTGACATTCTCCAGAG 3722
Db	
QY 772	CGCTGCCTGGGACTGAAACCCACTTTGGACTCTCGGCCCTGCCACAGTGACATTCTCCAGAG 831
Db	
QY 3723	AGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGCGTCTGGCGCTCGGCCCAAGG 3782
Db	
QY 832	AGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGCGTCTGGCGCTCGGCCCAAGG 891
Db	
QY 3783	ACTTCAACCGGCTCAAGAAAGATGCGAGCCTCAGGAGCTGCAGTCAACCTGCACTGG 3842
Db	
QY 892	ACTTCAACCGGCTCAAGAAAGATGCGAGCCTCAGGAGCTGCAGTCAACCTGCACTGG 951
Db	

QY 3843 GGGCTCATGGCTTCTGA 3859
|||
Db 952 GGGCTCATGGCTTCTGA 968

RESULT 15
AAK52532
ID AAK52532 standard; cDNA; 968 BP.

AAK52532;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 2061.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; KW
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

05-FEB-2001; 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 20-JUN-2000; 2000US-0598075.

PR 01-SEP-2000; 2000US-0654936.

PR 20-OCT-2000; 2000US-0693325.

XX
17710710 0000007 100007 401 00 177

[illegible]

PI Zhao QA, Wang D, Wang J,

[illegible]

DR P-PSDB; AAM79399.

PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -

PS Claim 1; Page 4448-4448; 6221pp; English.

CC The invention relates to polynucleotides

cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK32581), 2111 (AAK32582) and 3666 (AAK80920) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

Sequence 968 BP: 179 A: 353 C: 244 G: 192 T: 0 other:

Query Match	9.7%;	Score 495.4;	DB 22;	Length 968;
Best Local Similarity	99.8%;	Pred. No. 5.4e-113;		

Db	472	ATCTGAAC	TACCTGGGCC	CCCCCTTTCA	AGAGCAGACTTCA	ACCCTCCCGCCTGGGG	531
Qy	3423	CAGAGACT	TGCCAGG	CCACTGTTG	ACTTGGAGGT	TGGGGAAGCCTCA	TGACAAAC 3482
Db	532	CAGAGACT	TGCCAGG	CCACTGTTG	ACTTGGAGGT	TGGGGAAGCCTCA	TGACAAAC 591
Qy	3483	TGCGGCT	GACCCAG	AACACG	AGGCTTAC	AGCCACTTCTG	TGTTACTTGGCGTGGCTCA 3542
Db	592	TGCGGCT	GACCCAG	AACACG	AGGCTTAC	AGCCACTTCTG	TGTTACTTGGCGTGGCTCA 651
Qy	3543	ACCGTCA	GGCTGCC	ACTGCTG	AGCTGCG	CCGAGCCTTCTG	TGTTACTTGGCGTGGCTCC 3602
Db	652	ACCGTCA	GGCTGCC	ACTGCTG	AGCTGCG	CCGAGCCTTCTG	TGTTACTTGGCGTGGCTCC 711
Qy	3603	AGGGCT	TGCTGG	GCAGCAT	TGCGGG	CGTCATGGC	AGCTTGGGCTACCCACTGCCCCAGC 3662
Db	712	AGGGCT	TGCTGG	GCAGCAT	TGCGGG	CGTCATGGC	AGCTTGGGCTACCCACTGCCCCAGC 771
Qy	3663	CGCTGC	CTGGG	ACTGAA	CCACTT	GGACTCCTG	GGCCCTGCCACAGTGACTTCTCCAGA 3722
Db	772	CGCTGC	CTGGG	ACTGAA	CCACTT	GGACTCCTG	GGCCCTGCCACAGTGACTTCTCCAGA 831
Qy	3723	AGATGGA	CGACTT	CTGGCT	GTGA	AGGAGCTG	CAGACCTGGCTGTGGCGCTCGGCCAAGG 3782
Db	832	AGATGGA	CGACTT	CTGGCT	GTGA	AGGAGCTG	CAGACCTGGCTGTGGCGCTCGGCCAAGG 891
Qy	3783	ACTTCA	ACCGCT	CAAG	AGAGAT	GCAGCCT	CCAGCAGCTGCAGTCAACCCTGCACCTGG 3842
Db	892	ACTTCA	ACCGCT	CAAG	AGAGAT	GCAGCCT	CCAGCAGCTGCAGTCAACCCTGCACCTGG 951
Qy	3843	GGGCTC	ATGGCT	TCTGA	3859		
Db	952	GGGCTC	ATGGCT	TCTGA	968		

Search completed: February 1, 2003, 09:12:09
Job time : 1346.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 09:12:22 ; Search time 218.567 Seconds
(without alignments)
10456.327 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctggagtgggcctggc.....ccttgtaagtctctctca 5087

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5087	100.0	5087	10	US-09-931-704-3
2	523.4	10.3	797	10	US-09-931-704-1
C 3	493.4	9.7	495	10	US-09-864-761-23175
C 4	492	9.7	492	10	US-09-864-761-6462
5	437.2	8.6	465	10	US-09-867-701-4667
6	429.6	8.4	819	10	US-09-931-704-4
C 7	429.4	8.4	432	10	US-09-867-701-4636
C 8	355.8	7.0	360	10	US-09-867-701-2610
9	250.6	4.9	269	10	US-09-867-701-5382
10	192.8	3.8	231	10	US-09-867-701-2639
11	165.2	3.2	283	10	US-09-765-231A-75
12	57.6	1.1	53522	9	US-09-904-968A-1
13	54.4	1.1	1064	10	US-09-804-682-29
14	53	1.0	496	10	US-09-864-761-10689
15	53	1.0	1065	10	US-09-804-682-33
16	53	1.0	659158	9	US-09-771-208-20
17	52.4	1.0	267	10	US-09-864-761-27984
18	52.4	1.0	474	10	US-09-864-761-11284
C 19	52.4	1.0	3809	12	US-10-001-870-68

Sequence 20733, A
Sequence 3972, Ap
Sequence 74, Appl
Sequence 76, Appl
Sequence 554, App
Sequence 6203, App
Sequence 22817, A
Sequence 13, Appl
Sequence 17529, A
Sequence 19707, A
Sequence 2926, Ap
Sequence 27334, A
Sequence 22608, A
Sequence 5848, Ap
Sequence 8204, Ap
Sequence 3291, Ap
Sequence 6, Appl
Sequence 6, Appl
Sequence 1361, Ap
Sequence 18121, A
Sequence 1, Appl
Sequence 71, Appl
Sequence 73, Appl
Sequence 69, Appl
Sequence 106, App

US-09-864-761-20733
US-09-864-761-3972
US-10-101-487-74
US-10-101-487-76
US-09-954-456-554
US-09-864-761-6203
US-09-864-761-22817
US-09-870-962-13
US-09-864-761-17529
US-09-864-761-19707
US-09-864-761-2926
US-09-864-761-27334
US-09-864-761-22608
US-09-864-761-5848
US-09-864-761-8204
US-09-764-877-3291
US-09-924-400-6
US-09-810-936-6
US-09-429-755-6
US-09-864-761-1361
US-09-864-761-18121
US-10-014-717-1
US-10-101-487-71
US-10-101-487-73
US-10-101-487-69
US-10-101-487-106

52 1.0 575 10
52 1.0 1969 10
51.8 1.0 720 9
51.8 1.0 720 9
51.6 1.0 3301 10
49 1.0 475 10
49 1.0 511 10
48 0.9 1866 10
47.8 0.9 700 10
47.8 0.9 726 10
47.8 0.9 1951 10
47 0.9 360 10
46.8 0.9 272 10
46.8 0.9 454 10
46.4 0.9 600 10
45.4 0.9 32189 10
45.2 0.9 950 9
45.2 0.9 950 10
44.8 0.9 950 10
44.8 0.9 475 10
44.8 0.9 512 10
44.8 0.9 68750 9
44.6 0.9 522 9
44.6 0.9 530 9
44.6 0.9 554 9
44.6 0.9 554 9

ALIGNMENTS

RESULT 1

US-09-931-704-3
; Sequence 3, Application US/099311704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using
; TITLE OF INVENTION: Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 5087
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (137)..(138)
; OTHER INFORMATION: product = "INTERVENING UNSEQUENCED REGION OF >1KB"

Query Match 100.0%; Score 5087; DB 10; Length 5087;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACCTGGAGTGGGCTCGCGGATGGGATTATTAAGCTTCGCGGAGCGGGCTCGCC 60
Db 1 AACCTGGAGTGGGCTCGCGGATGGGATTATTAAGCTTCGCGGAGCGGGCTCGCC 60

Qy 61 CTCCTCCCTCCGAGCTCCCGGAGAGGAGCGGACCGCGGCGGCGCCAGCCGCCA 120
Db 61 CTCCTCCCTCCGAGCTCCCGGAGAGGAGCGGACCGCGGCGGCGCCAGCCGCCA 120

Qy 121 TGGACCTCCGAGAGGTTGAAAAACCCAACTAGCCCTCTTCAATACATGACAGCAG 180
Db 121 TGGACCTCCGAGAGGTTGAAAAACCCAACTAGCCCTCTTCAATACATGACAGCAG 180

Qy 181 CGCCCCATCTGATCACTAAACCGACCAAGTCACAGCCCTCCCACTTCCCTCTGCTGCC 240

Db 181 CGCCCCATCTGATACCTAAACCGACCAAGTCAAGCCCTCCAACTCACCTCTGCGTCC 240
Qy
Db 241 CAGACCTCACCATCTCTTGTGGAGCTCAAACTCAACCGCACTAAATCAACCAATCCCA 300
Db 241 CAGACCTCACCATCTCTTGTGGAGCTCAAACTCAACCGCACTAAATCAACCAATCCCA 300
Qy 301 AGTCTAAACTAACTGAACTTTTAAAGTAACCCAGTCTTAAACCTTAACCTAGCCCAAT 360
Db 301 AGTCTAAACTAACTGAACTTTTAAAGTAACCCAGTCTTAAACCTTAACCTAGCCCAAT 360
Qy 361 GCCAATTATATACCTAGCCAAACCCCTAACTGCTTTGCCAGTCCAAAGTGTCCACTG 420
Db 361 GCCAATTATATACCTAGCCAAACCCCTAACTGCTTTGCCAGTCCAAAGTGTCCACTG 420
Qy 421 AATCTCAGCTTGTGCTCACTGAAATCCGAAAGATATTTCCCACTGCCACAT 480
Db 421 AATCTCAGCTTGTGCTCACTGAAATCCGAAAGATATTTCCCACTGCCACAT 480
Qy 481 CCCTCTTACAGACCCCAACCTGGCTCTGGACTCTGGTATCTCTGGATGTCCTAACT 540
Db 481 CCCTCTTACAGACCCCAACCTGGCTCTGGACTCTGGTATCTCTGGATGTCCTAACT 540
Qy 541 CTGCACTGTCATCAGCCAAACAGCCGACTCGTCAAAATGCACTCTCTCCCTTCTGTCC 600
Db 541 CTGCACTGTCATCAGCCAAACAGCCGACTCGTCAAAATGCACTCTCTCCCTTCTGTCC 600
Qy 601 CCACTCTGAGCTGATGGAAGGGCTCATTTGAAGTCAACTTTTCCCACTAACACC 660
Db 601 CCACTCTGAGCTGATGGAAGGGCTCATTTGAAGTCAACTTTTCCCACTAACACC 660
Qy 661 AAGAACGGGGTGAACCTCCACACTGACCGTTCCTGAGAGTGAAGTCAACTTCTCTT 720
Db 661 AAGAACGGGGTGAACCTCCACACTGACCGTTCCTGAGAGTGAAGTCAACTTCTCTT 720
Qy 721 CAATCTAACCCCAACCTTACACTGCAAGTCACTAGAGTGAAGTCAACTTCTCTT 780
Db 721 CAATCTAACCCCAACCTTACACTGCAAGTCACTAGAGTGAAGTCAACTTCTCTT 780
Qy 781 ACTAGCCCCATAGGAGCGGAGCTAGTGTCTTAAACCTTACCTTCTCTTCTTCTT 840
Db 781 ACTAGCCCCATAGGAGCGGAGCTAGTGTCTTAAACCTTACCTTCTCTTCTTCTT 840
Qy 841 GTGAGTCTGTCTTGGGGGGGGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCT 900
Db 841 GTGAGTCTGTCTTGGGGGGGGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCT 900
Qy 901 AGCTGCCAGCTGTGATGTGTCTCCACCTCTGACTCTCCCTCAAGCTGCACTGCTC 960
Db 901 AGCTGCCAGCTGTGATGTGTCTCCACCTCTGACTCTCCCTCAAGCTGCACTGCTC 960
Qy 961 GACTGGAAGACTGGCAGGAGCTAGGTCAACTGGAACTAGGAGCTGCACTGCACTG 1020
Db 961 GACTGGAAGACTGGCAGGAGCTAGGTCAACTGGAACTAGGAGCTGCACTGCACTG 1020
Qy 1021 CCTAGGCTTGGCCCGCT 1080
Db 1021 CCTAGGCTTGGCCCGCT 1080
Qy 1081 GCACATGCGCAAGACTCTCTGAGTGAACACAGATCACTTCTCAAGTATCTACTGAT 1140
Db 1081 GCACATGCGCAAGACTCTCTGAGTGAACACAGATCACTTCTCAAGTATCTACTGAT 1140
Qy 1141 AGACACTCATGCTGCGCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
Db 1141 AGACACTCATGCTGCGCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
Qy 1201 GTCTTGCAGGAGTGTCT 1260
Db 1201 GTCTTGCAGGAGTGTCT 1260
Qy 1261 TGACCCCGACCCAGCCCAAGTGGGACAGACACTGAGGGGTGCGAGTGTCTCTCTCTCT 1320

Db 1261 TCACCCCCACCCAGCCCAAGTGGGACAGACACCTGAGGGGTGCCAGCTGTCTCTCC 1320
Qy 1321 GTGTGGCCCGGGCCGGCTCATGCTTCTGCTTCCATCTGCTCCACAGGGGACTGTGTTGG 1380
Db 1321 GTGTGGCCCGGGCCGGCTCATGCTTCTGCTTCCATCTGCTCCACAGGGGACTGTGTTGG 1380
Qy 1381 GATGTTAGGCTGTGTGCAAGTGTCTGGGACCTCTGCTGAGTGCAGTGTCTCAATCG 1440
Db 1381 GATGTTAGGCTGTGTGCAAGTGTCTGGGACCTCTGCTGAGTGCAGTGTCTCAATCG 1440
Qy 1441 CACAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTTACCTTACCCGCTACT 1500
Db 1441 CACAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTTACCTTACCCGCTACT 1500
Qy 1501 GGAGACCAACTCCGACAGTGTGGGACCTATGTGATATCCAGCGTAGGAACTCTGG 1560
Db 1501 GGAGACCAACTCCGACAGTGTGGGACCTATGTGATATCCAGCGTAGGAACTCTGG 1560
Qy 1561 AGTTGGGAGGAGTGTGGGAAAGACAGTCTCTAAACCTGAGGAGTGTCTGGTAA 1620
Db 1561 AGTTGGGAGGAGTGTGGGAAAGACAGTCTCTAAACCTGAGGAGTGTCTGGTAA 1620
Qy 1621 TGATGGGTGAGGAGGGCTCTTGGCTCCACAGTCCCTGCTCTCTCTCTCTCTCTCT 1680
Db 1621 TGATGGGTGAGGAGGGCTCTTGGCTCCACAGTCCCTGCTCTCTCTCTCTCTCTCT 1680
Qy 1681 CCCTTCTCTTCTAGTGGCCCTCCCTTCCCTTCCCTGCTCTCTCTCTCTCTCTCTCT 1740
Db 1681 CCCTTCTCTTCTAGTGGCCCTCCCTTCCCTTCCCTGCTCTCTCTCTCTCTCTCTCT 1740
Qy 1741 GGAGGCTCTGCAACCCGCTTGGGCACTGCTGCTTCCCTGCTCTCTCTCTCTCTCTCT 1800
Db 1741 GGAGGCTCTGCAACCCGCTTGGGCACTGCTGCTTCCCTGCTCTCTCTCTCTCTCT 1800
Qy 1801 TCCCTCTGGGGCCGGGAGTCTCTCTCTTACACCGTGTGTGTCTCTCTCTCTCTCT 1860
Db 1801 TCCCTCTGGGGCCGGGAGTCTCTCTCTTACACCGTGTGTGTCTCTCTCTCTCTCT 1860
Qy 1861 GGCGGGGTGGGTGGGACAGAGGGGCCCTCCCTGCTCTCTCTCTCTCTCTCTCTCT 1920
Db 1861 GGCGGGGTGGGTGGGACAGAGGGGCCCTCCCTGCTCTCTCTCTCTCTCTCTCTCT 1920
Qy 1921 TGCCCCAGACTGGGGCTCTGCTCTGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCT 1980
Db 1921 TGCCCCAGACTGGGGCTCTGCTCTGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCT 1980
Qy 1981 CCCATCTAGCTGGGCTCTAGGGGGTCTAGGGGAGGGGACTGTAGGGAACCCAGG 2040
Db 1981 CCCATCTAGCTGGGCTCTAGGGGGTCTAGGGGAGGGGACTGTAGGGAACCCAGG 2040
Qy 2041 CAGTAGTGGCAGGGGTTAGGGTGTGATGGAGTTATGTCTTAAGGATTTGGGGGTGG 2100
Db 2041 CAGTAGTGGCAGGGGTTAGGGTGTGATGGAGTTATGTCTTAAGGATTTGGGGGTGG 2100
Qy 2101 TCCAGAGTGTTCAGAGAGCCAGAGAGAGGAGGGGTTGGAGGAGCCAGGACCC 2160
Db 2101 TCCAGAGTGTTCAGAGAGCCAGAGAGAGGAGGGGTTGGAGGAGCCAGGACCC 2160
Qy 2161 ATGGGAAACCGGCCCT 2220
Db 2161 ATGGGAAACCGGCCCT 2220
Qy 2221 CCAGGAAAGAAAGGGGAGAGGTTGGGGGGAGCTGCTCAGCCCCAGGATACACCG 2280
Db 2221 CCAGGAAAGAAAGGGGAGAGGTTGGGGGGAGCTGCTCAGCCCCAGGATACACCG 2280
Qy 2281 AGGAAATTTAGTTGTCTCTGCTGTTGTCAGCTGTGAACCTCCCTCTGGGCTTGGCTA 2340
Db 2281 AGGAAATTTAGTTGTCTCTGCTGTTGTCAGCTGTGAACCTCCCTCTGGGCTTGGCTA 2340
Qy 2341 TCCAGGCT 2400
Db 2341 TCCAGGCT 2400

[illegible]

QY	3481	ACTGGGCTGACCCAGAACTTACGAGCGCTACAGCCACCTTCTGTGTATTCTGTGGTGGCGCT	3540
DB	3481	ACTGGGCTGACCCAGAACTACGAGCGCTACAGCCACCTTCTGTGTATTCTGTGGTGGCGCT	3540
QY	3541	CAACCGTCAGGCTGCCACTGCTGAGCTGCGCGCAGCGCTGGCCACTTCTGCAACCAAGCCT	3600
DB	3541	CAACCGTCAGGCTGCCACTGCTGAGCTGCGCGCAGCGCTGGCCACTTCTGCAACCAAGCCT	3600
QY	3601	CCAGGCGCTGCTGGGCGAGCATTTGCGGCGCTCATGGCAGCTCTGGGCTACCCATGCGCCCA	3660
DB	3601	CCAGGCGCTGCTGGGCGAGCATTTGCGGCGCTCATGGCAGCTCTGGGCTACCCATGCGCCCA	3660
QY	3661	GCGCGTCCCTGGGACTGAACCCACTTTGGAATCTCTGGCCCTGCCACAGTGACTTCTCTCA	3720
DB	3661	GCGCGTCCCTGGGACTGAACCCACTTTGGAATCTCTGGCCCTGCCACAGTGACTTCTCTCA	3720
QY	3721	GAAGATGGACGACTTCTGCTGCTGAAGAGCTGACAGACTGCTGTGGCGCTCGGCCAA	3780
DB	3721	GAAGATGGACGACTTCTGCTGCTGAAGAGCTGACAGACTGCTGTGGCGCTCGGCCAA	3780
QY	3781	GGACTTCAACCGGCTCAAGAAGAATGCAAGCTCCAGCAGCTTGCACTCACCTTGCACTT	3840
DB	3781	GGACTTCAACCGGCTCAAGAAGAATGCAAGCTCCAGCAGCTTGCACTCACCTTGCACTT	3840
QY	3841	GGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCCCTTCAAAACCTGTCT	3900
DB	3841	GGGGCTCATGGCTTCTGACTTCTGACTTCTGACTTCTGCTTCTGCTTCCCTTCAAAACCTGTCT	3900
QY	3901	CCCACTTTGTGAGAGCCAGCCCTGTATGCAACACCTGTTGAGCCAGGAGACAGAAGCTG	3960
DB	3901	CCCACTTTGTGAGAGCCAGCCCTGTATGCAACACCTGTTGAGCCAGGAGACAGAAGCTG	3960
QY	3961	TGAGCTCTTGGGCCCTTCTTGACCGGCTGGGCGTGTGATCGCATCAGCCCTGTCTCTCTC	4020
DB	3961	TGAGCTCTTGGGCCCTTCTTGACCGGCTGGGCGTGTGATCGCATCAGCCCTGTCTCTCTC	4020
QY	4021	CCCACTCCCAAAGTCTACCAAGCTGGGAGGAGTACAGTAGGCGCTGTCTGTCTCTG	4080
DB	4021	CCCACTCCCAAAGTCTACCAAGCTGGGAGGAGTACAGTAGGCGCTGTCTGTCTCTG	4080
QY	4081	TTTCTACAGGAAGTCATGCTCAGGAGGTGTGAAGTGGTTTCTGTTGGTGCAGAGGGCT	4140
DB	4081	TTTCTACAGGAAGTCATGCTCAGGAGGTGTGAAGTGGTTTCTGTTGGTGCAGAGGGCT	4140
QY	4141	CATGGCTCTCTCTTCTGCTTACCACTTGGCCAGTGCCACCCAGCCCTCAGGTGGCA	4200
DB	4141	CATGGCTCTCTCTTCTGCTTACCACTTGGCCAGTGCCACCCAGCCCTCAGGTGGCA	4200
QY	4201	CATCTGAGGGCAGGGTTGAGGGCCACCAACACATGCTTCTGGGGTGAAGCCCT	4260
DB	4201	CATCTGAGGGCAGGGTTGAGGGCCACCAACACATGCTTCTGGGGTGAAGCCCT	4260
QY	4261	TTGGCTGCCCACTCTCTTGAATGGGTGTGCTTATCCCAAAATCAGTCTATACA	4320
DB	4261	TTGGCTGCCCACTCTCTTGAATGGGTGTGCTTATCCCAAAATCAGTCTATACA	4320
QY	4321	TCCAATTCAGGAAACAAACATCGTGGCAATCTTACACAAAAGAGATGAGATTAAAGTG	4380
DB	4321	TCCAATTCAGGAAACAAACATCGTGGCAATCTTACACAAAAGAGATGAGATTAAAGTG	4380
QY	4381	CAGGGTTGGGGTCTGCAATTTGAGGTGCCCTATAAAACCAAGAGAAAATACTGAAAGCAC	4440
DB	4381	CAGGGTTGGGGTCTGCAATTTGAGGTGCCCTATAAAACCAAGAGAAAATACTGAAAGCAC	4440
QY	4441	AGGGCAGGGAACAGACAGACCCAGAGTCTTCAAAGCACAGGTGGGCAAAACAA	4500
DB	4441	AGGGCAGGGAACAGACAGACCCAGAGTCTTCAAAGCACAGGTGGGCAAAACAA	4500
QY	4501	ACCCGAGTGTAGCATCAGGACCTTGTCTTCCAGTATTTACGGTGCCTCTTC	4560
DB	4501	ACCCGAGTGTAGCATCAGGACCTTGTCTTCCAGTATTTACGGTGCCTCTTC	4560
QY	4561	TCTGCCCTCTTTCCCAAGGTATCTGTGGGTGCGCAGCTGGGGAGGGCAACATAGCCAC	4620


```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23175
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALU8 8.00e-03
; OTHER INFORMATION: NT HIT: G114439486, EVALU8 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: A1752561.1, EVALU8 0.00e+00
US-09-864-761-23175

```

Query Match	9.7%; Score 493.4; DB 10; Length 495;
Best Local Similarity	99.8%; Pred. No. 1.7e-122;
Matches 494; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY 3365	CTGAACCTACCTGGGCCCCCTTTCAAAGAGCCAGACTTCAACCTTCCCGCTCGGGGCA 3424
Db	
QY 495	CTGAACCTACCTGGGCCCCCTTTCAAAGAGCCAGACTTCAACCTTCCCGCTCGGGGCA 436
Db	
QY 3425	GAGACTCTGCCCCAGGGCCACTGTTGACTTGGAGGTGGCGAAGCCTCAATGACAAACTG 3484
Db	
QY 435	AAGACTCTGCCCCAGGGCCACTGTTGACTTGGAGGTGGCGAAGCCTCAATGACAAACTG 376
Db	
QY 3485	CGGCTGACCCAGAACTAGAGGCCCTACAGGCACCTTCTGTGTTACTTTCGTGGCTCAAC 3544
Db	
QY 375	CGGCTGACCCAGAACTAGAGGCCCTACAGGCACCTTCTGTGTTACTTTCGTGGCTCAAC 316
Db	
QY 3545	CGTCAGGCTGCACCTGCTGAGCTGCGCGGAGCCTTGGCCACCTTCTGACACAGCCTCCAG 3604
Db	
QY 315	CGTCAGGCTGCCACTGCTGAGCTGCGCGGAGCCTTGGCCACCTTCTGACACAGCCTCCAG 256
Db	
QY 3605	GGCCTGCTGGCGCAGCTTGGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGCCG 3664
Db	
QY 255	GGCCTGCTGGCGCAGCTTGGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGCCG 196
Db	
QY 3665	CTGCCTGGGACTGAAACCCACTTGGACTCTGGCCCTGCCACAGTGACTTCTCCCGAAG 3724
Db	
QY 195	CTGCCTGGGACTGAAACCCACTTGGACTCTGGCCCTGCCACAGTGACTTCTCCCGAAG 136
Db	
QY 3725	ATGGACGACTTTCGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGGAC 3784
Db	
QY 135	ATGGACGACTTTCGGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGGAC 76
Db	
QY 3785	TTCAACCGGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTGCAGTACCCCTGCACTCTGGG 3844
Db	
QY 75	TTCAACCGGCTCAAGAAGAAGATGCAGCCTCCAGCAGCTGCAGTACCCCTGCACTCTGGG 3844
Db	
QY 3845	GCTCATGGCTTTCGA 3859
Db	
QY 15	GCTCATGGCTTTCGA 1
Db	

RESULT 4

US-09-864-761-6462/c

```

/ Sequence 6462, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharron G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC
/ TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
/ FILE REFERENCE: Aeonica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/006666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006657
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006654
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006659
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006655
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006658
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006653
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006652
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006651
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/006670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 6462
/ LENGTH: 492
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005849.1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
/ OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.2
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
/ US-09-864-761-6462

```

Query Match 9.7%; Score 492; DB 10; Length 492;
Best Local Similarity 100.0%; Pred.No. 4.1e-122;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	3519	TTCTGTGTTACTTGCCTGGCTCAACCGTCAGGCTGCCACTGCTAGCTGC	CGCCGAGCC	3578
Dβ	492	TTCTGTGTTACTTGCCTGGCTCAACCGTCAGGCTGCCACTGCTAGCTGC	CGCCGAGCC	433

Qy	3603	AGGGCTGCTGGGAGCATTTGGGGGGGTATGCGAGCTCTGGGCTACCCACTGCCCCCAGC	3662
Db	516	AGGGCTGCTGGGAGCATTTGCGAGTGTATGCGCAGCGTTGGCTACCCACTGCCCCCAGC	575
Qy	3663	CGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTGACTTCTCCAGA	3722
Db	576	CTCTGCCCAGGACATGAGCCAGCGCTGGGCCCTTGGCCCTGCCACAGTGACTTCTCCAGA	635
Qy	3723	AGATGGACGACTTCTTGGCTGCTGAAGGAGTGTGCAGACCTTGGCTGTGGGGCTCGGCCAAGG	3782
Db	636	AGATGGATGACTTCTTGGCTGCTGAAGGAGTGTGCAGACCTTGGCTGTGGGGCTCGGCCAAGG	695
Qy	3783	ACTTTCACCCGGCTCAAGAGAAATGCGAGCTCCAGCAGCTGCGAGTCACTCCGACCTGG	3842
Db	696	ACTTTCACCCGGCTTAAAGAGAAATGCGAGCTCCAGCAGCTTCACTCACTCCGACCTGG	755
Qy	3843	GGGGCTCATGGCTTCTGACTTCTTGACCTT	3870
Db	756	AGGCACATGGTTTCTGACCTCTGACCT	783

RESULT 7

US-09-867-701-4636/c

; Sequence 4636, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Agiate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4636

; LENGTH: 432

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-867-701-4636

QY	853	TTGCGGCCGC	863
Db	12	TTGCGGCCGC	2
RESULT 8			
US-09-867-701-2610/c			
; Sequence 2610, Application US/09867701			
; Patent No. US20020132237A1			
; GENERAL INFORMATION:			
; APPLICANT: Aglate, Paul A.			
; APPLICANT: Jones, Robert			
; APPLICANT: Harlocker, Susan L.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY			
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER			
; FILE REFERENCE: 210121.497			
; CURRENT APPLICATION NUMBER: US/09/867,701			
; CURRENT FILING DATE: 2001-05-29			
; NUMBER OF SEQ ID NOS: 10912			
; SOFTWARE: Fast-SEQ for Windows Version 4.0			
; SEQ ID NO 2610			
; LENGTH: 360			
; TYPE: DNA			
; ORGANISM: Homo sapien			
US-09-867-701-2610			
Query Match 7.0%; Score 355.8; DB 10; Length 360;			
Best Local Similarity 99.4%; Pred. No. 1.3e-85;			
Matches 357; Conservative 0; Mismatches 2; Indels 0; Gaps			
QY	505	GCCTCTGGACTCTGGTATCCTGGGATGTCCAAACTCTGCAGTGCCATCAGGCCAACCAAGC	564
Db	360	GCCTCTGGACTCTGGTATCCTGGGATGTCCAAACTCTGCAGTGCCATCAGGCCAACCAAGC	301
QY	565	CCGACTCGTCAAATGCACCTCTCTCCCTTCCTGTGCCACCCTTGCAGGCTGATGGAAG	624
Db	300	CCGACTCGTCAAATGCACCTCTCTCCCTTCCTGTGCCACCCTTGCAGGCTGATGGAAG	241
QY	625	GCCTCATTTGAAGTCCAACCTTTTCCCCCACCTTAACACCAAGAAGGGGTGAACCTCCACACT	684
Db	240	GCCTCATTTGAAGTCCAACTTTTCCCCCACCTTAACACCAAGAAGGGGTGAACCGCCACACT	181
QY	685	GCCACCGTTCCCTTGAGAGTGAGCACTAAATCTCTCTTCAATCTAAACCCCAACCTTACACTTC	744
Db	180	GCCACCGTTCCCTTGAGAGTGAGCACTAAATCTCTCTTCAATCTAAACCCCAACCTTACACTTC	121
QY	745	CCACACTCAGGAATCACATCCTAGAATAATACCCAAAACCTAAGCCCCTAAGCGACGCCGA	804
Db	120	CCACACTCAGGAATCACATCCTAGAATAATACCCAAAACCTAAGCCCCTAAGCGACGCCGA	61
QY	805	CCCTAGTGGTCTAACCCCTATACCTTGCTTCCTATGGGTGAGTCTGTTCTTGGCGGCGCG	863
Db	60	CCCTAGTGGTCTAACCCCTATACCTTGCTTCCTATGGGTGAGTCTGTTCTTGGCGGCGCG	2

Query Match	8.4%; Score 429.4; DB 10; Length 432;	
Best Local Similarity	99.8%; Pred. No. 2.4e-105;	
Matches 430; Conservative	0; Mismatches 1; Indels 0; Gaps	0;
QY	433	GGTCCTCACTCAAAATCCAGAAAAGCATATTTCCCCACTGCCACACATCCCTCCTTACAG 492
Db	432	GGTCCTCACTTAAATCCAGAAAAGCATATTTCCCCACTGCCACACATCCCTCCTTACAG 373
QY	493	CACCCAAACCCCTGGCCTCTGGACTCTCGTATCTCTGGGATGTCCAAACTCTCGACGTGCCAT 552
Db	372	CACCCAAACCCCTGGCCTCTGGACTCTCGTATCTCTGGGATGTCCAAACTCTCGACGTGCCAT 313
QY	553	CAGCCAAACAAGCCGACTCGTCAATGACACTCTCTCCCTTCCTGTCCCAACCTTTGGAG 612
Db	312	CAGCCAAACAAGCCGACTCGTCAATGACACTCTCTCCCTTCCTGTCCCAACCTTTGGAG 253
QY	613	GCTGATGGAAAGGCCTCATTTGAAGTCCAACTTTTCCCAACCTTAACACCAAGAACGGGGTG 672
Db	252	GCTGATGGAAAGGCCTCATTTGAAGTCCAACTTTTCCCAACCTTAACACCAAGAACGGGGTG 193
QY	673	AACCTCCACACTGCACCGTTCCCTGAGGTGACACTAAATCTCCTTCAATCTAACCC 732
Db	192	AACCTCCACACTGCACCGTTCCCTGAGGTGACACTAAATCTCCTTCAATCTAACCC 133
QY	733	ACCTTACACTTCCACACTCAGGAATCACATCTCTAGAATATATCCCAAACTTAAGCCCCAT 792
Db	132	ACCTTACACTTCCACACTCAGGAATCACATCTCTAGAATATATCCCAAACTTAAGCCCCAT 73
QY	793	AAGCGAGCCGACCTAGTGGTCTTAACCCCTATACCTTGTCTCTATGGGTGAGTCTGTT 852
Db	72	AAGCGAGCCGACCTAGTGGTCTTAACCCCTATACCTTGTCTCTATGGGTGAGTCTGTT 13

Db 34289 CCTCTCTCCCTCCCTAGCCCTTCCCTCTTCCCCCCAGCCCTTCCCCCTCTCCCCC 34348

[illegible]

RESULT 13

```

US-09-804-682-29
; Sequence 29, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1064
; TYPE: DNA
; ORGANISM: Homo sapiens

```

RESULT 14

US-09-864-761-10689
; Sequence 10689, Application US/09864761
; Patent No. US20020048763A1


```
Db 433 -----CTTCCGCCGCCCTCTCTCNCCGCCGCCCTTCCCTCCCTCNCCCTCNCAACCCCC 487
Qy 2572 ACCCGTCTCTCCCGAGCCCTCAGCTGGTGGGCTTGGGTGTGTAGCGGCAAAATGGGGCT 2631
Db 488 TTCCCCNTCCGCCCTCTCTCNCCGCCGCCCTTCCCTCCCTCNCCCTCNCAACCCCC 547
Qy 2632 CTGGTTCGAATGGGGCACTCTCATCTCTCTCTTGTCTTGTGTGAGAAAACTTTGCTTC 2691
Db 548 CCCCCCCCCCTCTCTCCGCCGCCCTTCCCTCCCTCNCCGCCGCCCTTCCCTCCCTCNCCGCCGCC 607
Qy 2692 ACTCCACTGCGCTCTCTAGTTCGAGCCCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2751
Db 608 CCCCCCCCCCTCTCNCCGCCGCCCTTCCCTCCCTCNCCGCCGCCCTTCCCTCCCTCNCCGCCGCC 667
Qy 2752 CTCGAAGGAGTGTCTACACCTCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2811
Db 668 CCNCGTTCCCAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 727
Qy 2812 CCCCCTGCAATCTGGGTTCCAGGCCGCCAGCAATGGTTCTCTCAAGGTCTCAGGACCTC 2871
Db 728 CCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 787
Qy 2872 CTTCGCAAGC-CCGACAGTGTGTTGAAGGCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2930
Db 788 CCTCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 847
Qy 2931 CACTGCTGAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2990
Db 848 CCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 907
Qy 2991 CCACCTTCTACTCTCTCAGCTCTCTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3050
Db 908 CCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 967
Qy 3051 GGGCACTCTCCCAAGGTTTGGCCAGCCAGCAATCAGCAAGTCTCTCTCTCTCTCTCTCTCTCT 3110
Db 968 TCCCNCTCCCCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1027
Qy 3111 GCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3144
Db 1028 NNCTTCNCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1061
```

Search completed: February 2, 2003, 03:58:46
Job time : 412.567 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 163.925 Seconds
(without alignments)
9516.918 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 5087

Sequence: 1 aacctgcgagtggcctggc.....cctttgtaagtcttctca 5087

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5087	100.0	5087	1	US-08-792-019B-3
2	5087	100.0	5087	3	US-08-988-819-3
3	5087	100.0	5087	3	US-08-016-534-3
4	1463.4	28.8	1710	3	US-08-106-182-1
5	523.4	10.3	797	1	US-08-792-019B-1
6	523.4	10.3	797	3	US-08-988-819-1
7	523.4	10.3	797	3	US-08-016-534-1
8	429.6	8.4	819	1	US-08-792-019B-4
9	429.6	8.4	819	3	US-08-988-819-4
10	429.6	8.4	819	3	US-08-016-534-4
11	324.6	6.4	396	3	US-08-106-182-7
12	99	1.9	7218	1	US-08-232-463-14
13	60.2	1.2	1926	4	US-08-249-585A-4
14	60.2	1.2	1931	2	US-08-130-114-2
15	57.6	1.1	53526	3	US-08-658-136-2
16	57.6	1.1	53577	3	US-08-658-136-1
17	52.4	1.0	51259	3	US-08-781-891-209
18	52.2	1.0	1166	4	US-08-072-596-323
19	49.6	1.0	16442	3	US-08-781-891-208
20	48.8	1.0	1926	4	US-08-249-585A-2
21	48.8	1.0	2580	3	US-08-050-863-2
22	48.8	1.0	2580	4	US-08-359-081-2
23	48.8	1.0	5452	2	US-08-130-114-1
24	48.8	1.0	7218	1	US-08-232-463-14
25	48.8	1.0	9600	4	US-08-910-647-1
26	48.8	1.0	9600	4	US-08-620-525-1
27	48.8	1.0	10596	1	US-07-884-811-15

C 28	48.8	1.0	10596	1	US-07-885-971-15	Sequence 15, Appl
C 29	48.8	1.0	10596	1	US-08-087-783A-15	Sequence 15, Appl
C 30	48.8	1.0	10596	1	US-08-194-088B-15	Sequence 15, Appl
C 31	48.8	1.0	10596	2	US-08-194-087-15	Sequence 15, Appl
C 32	48.8	1.0	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 33	48.6	1.0	289	4	US-09-007-005-17	Sequence 17, Appl
C 34	48.6	1.0	289	4	US-09-244-796-17	Sequence 17, Appl
C 35	48	0.9	1866	3	US-09-173-581-13	Sequence 13, Appl
C 36	48	0.9	1866	4	US-09-420-915-13	Sequence 13, Appl
C 37	45.2	0.9	950	4	US-08-991-789A-6	Sequence 6, Appl
C 38	45.2	0.9	950	4	US-08-062-451-6	Sequence 6, Appl
C 39	45.2	0.9	950	4	US-09-598-326-6	Sequence 6, Appl
C 40	44.8	0.9	68750	3	US-09-335-409-1	Sequence 1, Appl
C 41	44.8	0.9	68750	4	US-09-568-102-1	Sequence 1, Appl
C 42	44.8	0.9	68750	4	US-09-567-969-1	Sequence 1, Appl
C 43	44.8	0.9	68750	4	US-09-568-480-1	Sequence 1, Appl
C 44	44.8	0.9	68750	4	US-09-568-486-1	Sequence 1, Appl
C 45	44.8	0.9	68750	4	US-09-568-472-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-792-019B-3
; Sequence 3, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED
; OTHER INFORMATION: REGION OF >1KB"
US-08-792-019B-3

Query Match 100.0%; Score 5087; DB 1; Length 5087;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5087; Conservative 0; Mismatches 0; Gaps 0;
Qy 1 AACCTGGAGTGGCGCTGGCGGATGGGATTTAAAGCTTCGGGAGCCGGCTGCC 60
Db 1 AACCTGGAGTGGCGCTGGCGGATGGGATTTAAAGCTTCGGGAGCCGGCTGCC 60

QY 61 CTCCCACTCCGAGCCTCCGGAGAGGAGCGCACCCCGCGCGCCAGCCCGAGCCCA 120
Db 61 CTCCCACTCCGAGCCTCCGGAGAGGAGCGCACCCCGCGCGCCAGCCCGAGCCCA 120
QY 121 TGGACCTCCGAGCAGGTTGAAAACCCAAACTAGCCCTGCTCTTTATAAATGAAAGCAG 180
Db 121 TGGACCTCCGAGCAGGTTGAAAACCCAAACTAGCCCTGCTCTTTATAAATGAAAGCAG 180
QY 181 CGCCCATCTGATACCTAAACCGACCAAGTCACAGCGCTCCAACTCACCTCTGCGCTGC 240
Db 181 CGCCCATCTGATACCTAAACCGACCAAGTCACAGCGCTCCAACTCACCTCTGCGCTGC 240
QY 241 CAGACCTCACCATCTCTTTGTGAGCTCAAACTCAACCGCACTAAATCAACCAAAATCCCA 300
Db 241 CAGACCTCACCATCTCTTTGTGAGCTCAAACTCAACCGCACTAAATCAACCAAAATCCCA 300
QY 301 AGTCTAAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTTAAACCTTAACTAGCCCAAT 360
Db 301 AGTCTAAACTAATCTGAAACTTTTAAAGTAAACCCAGTCTTTAAACCTTAACTAGCCCAAT 360
QY 361 GCGAATTATATCTACCTAGCGAAACCGCTAACTGCGCTTTGCCAGTCCAAAGTGTCCACTG 420
Db 361 GCGAATTATATCTACCTAGCGAAACCGCTAACTGCGCTTTGCCAGTCCAAAGTGTCCACTG 420
QY 421 AATCTCTACCTGGTCTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCCCACAT 480
Db 421 AATCTCTACCTGGTCTCTCACTGAAATCCAGAAAGCATATTTCCCACTGCCCCACAT 480
QY 481 CCTCTCTTACAGCACCCACTGGCTCTGGA CTCTCTGGTATCTCTGGGATGTCCAAACT 540
Db 481 CCTCTCTTACAGCACCCACTGGCTCTGGA CTCTCTGGTATCTCTGGGATGTCCAAACT 540
QY 541 CTGCGAGTGCATCAGCGAAACCGCTCGTCAAAATGCACTCTCTCCCTCTCTCTGTC 600
Db 541 CTGCGAGTGCATCAGCGAAACCGCTCGTCAAAATGCACTCTCTCCCTCTCTCTGTC 600
QY 601 CCACCTCTGAGGCTGATGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAAACAC 660
Db 601 CCACCTCTGAGGCTGATGAAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAAACAC 660
QY 661 AAGAACGGGTGAACCTTCCA CTCTGCAACCGTTCCCTGAGAGTGAGCACTTAAATCTCCTT 720
Db 661 AAGAACGGGTGAACCTTCCA CTCTGCAACCGTTCCCTGAGAGTGAGCACTTAAATCTCCTT 720
QY 721 CAATCTAACCCCACTTACATCTTCCACACTCAGGAATCACTCTAGAAATATACCCAAA 780
Db 721 CAATCTAACCCCACTTACATCTTCCACACTCAGGAATCACTCTAGAAATATACCCAAA 780
QY 781 ACTAAGCCCCATAGGAGCGAGCCGACCTAGTGTCTTAACTTATACCTTGTCTCTATGG 840
Db 781 ACTAAGCCCCATAGGAGCGAGCCGACCTAGTGTCTTAACTTATACCTTGTCTCTATGG 840
QY 841 GTGAGTCTGTTCTTGGCGGCGGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 841 GTGAGTCTGTTCTTGGCGGCGGCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 901 AGCCTGCACTCTGATGCTGTCTCTCCACCTCTGACTCCCTCAAGCTCAGTGG 960
Db 901 AGCCTGCACTCTGATGCTGTCTCTCCACCTCTGACTCCCTCAAGCTCAGTGG 960
QY 961 GACTGGAAGACTTGGCAGGAAGTATAGGTTAACTGGAACACAGGAGGTCCAGCTGCACT 1020
Db 961 GACTGGAAGACTTGGCAGGAAGTATAGGTTAACTGGAACACAGGAGGTCCAGCTGCACT 1020
QY 1021 CCTTAGGCTTGGCCCTGCT 1080
Db 1021 CCTTAGGCTTGGCCCTGCT 1080
QY 1081 GCACATGCGAAGACTCTCTAGCTGACACAGATCCATCTCTCAAGTATCTACTCAT 1140
Db 1081 GCACATGCGAAGACTCTCTAGCTGACACAGATCCATCTCTCAAGTATCTACTCAT 1140
QY 1141 AGACACTCATGCGTGCAGAGTCTCTCATCTCTCAAAACATACATATGCTCTCTCTCTCT 1200

Db 1141 AGACACTCATGCGTGCAGAGTCTCTCATCTCAAAACATACATATGCTCTCTCTCTCTCT 1200
QY 1201 GTCTTGGCAGAGAGTCTTCCCTCTCTCATCTCCATCTCCCTCCATCTGGTGTCCACCC 1260
Db 1201 GTCTTGGCAGAGAGTCTTCCCTCTCTCATCTCCCTCTCTCCATCTGGTGTCCACCC 1260
QY 1261 TCACCCCCCAACCCAGCCCAAGTGGGAGACACACCTGAGGGGTGCCAGTGTCTCTCC 1320
Db 1261 TCACCCCCCAACCCAGCCCAAGTGGGAGACACACCTGAGGGGTGCCAGTGTCTCTCC 1320
QY 1321 GTGTGGGCGGCGGCGGCTCATGCTTCTGTCATCTGTCATCTGCCCCACAGGGGACTCTGGGG 1380
Db 1321 GTGTGGGCGGCGGCGGCTCATGCTTCTGTCATCTGTCATCTGCCCCACAGGGGACTCTGGGG 1380
QY 1381 GATGTTAGCGTGTCTGTCACGCTCTCTGTCACCTCTCTGTCAGTGTCCAGTCTCTCAATCG 1440
Db 1381 GATGTTAGCGTGTCTGTCACGCTCTCTGTCACCTCTCTGTCAGTGTCCAGTCTCTCAATCG 1440
QY 1441 CACAGGGGACCCAGGGCTGCGCTCCATCCAGAAAACTTATGACCTCACCGCTACCT 1500
Db 1441 CACAGGGGACCCAGGGCTGCGCTCCATCCAGAAAACTTATGACCTCACCGCTACCT 1500
QY 1501 GGAGCAACACTCCGAGCTTGGCTGGGACCTATGTGAGTATCCAGCTGAGGAATCTGGG 1560
Db 1501 GGAGCAACACTCCGAGCTTGGCTGGGACCTATGTGAGTATCCAGCTGAGGAATCTGGG 1560
QY 1561 AGTTGGGAGGAGTGGGAGTTGGGAAAGACAGTCTTAAACCGTGGAGGGTCTCTGTGATAA 1620
Db 1561 AGTTGGGAGGAGTGGGAGTTGGGAAAGACAGTCTTAAACCGTGGAGGGTCTCTGTGATAA 1620
QY 1621 TGATGGGTTGAGAGGGGCTCTTTGGCTCCCAACAGTCTCTCTGTCTGTCTATCTCTG 1680
Db 1621 TGATGGGTTGAGAGGGGCTCTTTGGCTCCCAACAGTCTCTCTGTCTGTCTATCTCTG 1680
QY 1681 CCTTCTCTCTTGTAGTGGGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 1740
Db 1681 CCTTCTCTCTTGTAGTGGGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 1740
QY 1741 GGCGAGGCTCGCACCGGCTTGGGCGGCTTGGGCGGCTTGGGCGGCTTGGGCGGCTTGGG 1800
Db 1741 GGCGAGGCTCGCACCGGCTTGGGCGGCTTGGGCGGCTTGGGCGGCTTGGGCGGCTTGGG 1800
QY 1801 TCCCT 1860
Db 1801 TCCCT 1860
QY 1861 GGCGGGTGGTGGGAGACAGAGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920
Db 1861 GGCGGGTGGTGGGAGACAGAGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920
QY 1921 TGCCCCCAGACTGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1980
Db 1921 TGCCCCCAGACTGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1980
QY 1981 CCCATCTAGCTGGGCGGCT 2040
Db 1981 CCCATCTAGCTGGGCGGCT 2040
QY 2041 CAGTAGTGGCAGGGGTTTGGGTTGAGTGGAGTTATGCTGTAAAGATTTGGGGGTGG 2100
Db 2041 CAGTAGTGGCAGGGGTTTGGGTTGAGTGGAGTTATGCTGTAAAGATTTGGGGGTGG 2100
QY 2101 TCCAGAGTGTTCAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
Db 2101 TCCAGAGTGTTCAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
QY 2161 ATGGGGAACCGGCGGCT 2220
Db 2161 ATGGGGAACCGGCGGCT 2220
QY 2221 CCAGGGAAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280

Db 2221 CAGGGAAAGAAAGGAGAGAGGTGGCGGGGAGCTGGCTCAGCGCCAGGATACACCG 2280
Qy 2281 AGGAAATAGTTTGTCTCTGTCTGTGTGACGCTGTGAACTCTCCCTCGGCGCTTGCCTA 2340
Db 2281 AGGAAATAGTTTGTCTCTGTCTGTGTGACGCTGTGAACTCTCCCTCGGCGCTTGCCTA 2340
Qy 2341 TCCAGGCTCTCCCTTGTCT 2400
Db 2341 TCCAGGCTCTCCCTTGTCT 2400
Qy 2401 TCCCTGGGCCCCAGCGCTCCCCGAGGGTTGAAAGGGCTCTGCGCTCTCTCTCTCTCT 2460
Db 2401 TCCCTGGGCCCCAGCGCTCCCCGAGGGTTGAAAGGGCTCTGCGCTCTCTCTCTCTCT 2460
Qy 2461 ATGCTGTCTTCCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520
Db 2461 ATGCTGTCTTCCATAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520
Qy 2521 CTGCAACCTGTCTCTATCAGCTGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2580
Db 2521 CTGCAACCTGTCTCTATCAGCTGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2580
Qy 2581 TCCCAAGCCCCAGCTGGTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2640
Db 2581 TCCCAAGCCCCAGCTGGTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2640
Qy 2641 ATGGGCACT 2700
Db 2641 ATGGGCACT 2700
Qy 2701 CCTCTCTAGTTTCCGACCT 2760
Db 2701 CCTCTCTAGTTTCCGACCT 2760
Qy 2761 GTGGTCTACACCTCTGCT 2820
Db 2761 GTGGTCTACACCTCTGCT 2820
Qy 2821 TCTGGCTTCCAGGCTTCCAGCAATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
Db 2821 TCTGGCTTCCAGGCTTCCAGCAATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880
Qy 2881 CCCGACAGTTTGAAGCTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Db 2881 CCCGACAGTTTGAAGCTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Qy 2941 CGTCTGCT 3000
Db 2941 CGTCTGCT 3000
Qy 3001 CCTCTCCAGCT 3060
Db 3001 CCTCTCCAGCT 3060
Qy 3061 CCAAGTTTGGCCACCCAGCCAAATCAGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120
Db 3061 CCAAGTTTGGCCACCCAGCCAAATCAGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3120
Qy 3121 CT 3180
Db 3121 CT 3180
Qy 3181 ACCTGCATACAAATGATATCTTATGAAATCTAGGAGGCTCAGGAGGCTCAGGAGGCT 3240
Db 3181 ACCTGCATACAAATGATATCTTATGAAATCTAGGAGGCTCAGGAGGCTCAGGAGGCT 3240
Qy 3241 TAGCATGGAGACAGGGCT 3300
Db 3241 TAGCATGGAGACAGGGCT 3300
Qy 3301 TCTCTTGGTGGTGAAGGCT 3360
Db 3301 TCTCTTGGTGGTGAAGGCT 3360

Qy 3361 ACAGCTGAAGTACTCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCTGGG 3420
Db 3361 ACAGCTGAAGTACTCTGGGCCCCCTTTCAAGAGCCAGACTTCAACCCCTCCCGCTGGG 3420
Qy 3421 GGAGAGAGCTCTGCCCCAGGGCCACTCTTGAATTTGAGAGTGTGGAGAGCTCAATGACAA 3480
Db 3421 GGAGAGAGCTCTGCCCCAGGGCCACTCTTGAATTTGAGAGTGTGGAGAGCTCAATGACAA 3480
Qy 3481 ACTGCGGCTGAGCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGTGTGSCCT 3540
Db 3481 ACTGCGGCTGAGCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGTGTGSCCT 3540
Qy 3541 CAAACCTGAGGCTGCGGCTGAGCTGCGGCGCAGCTGCGGCTGCGGCTGCGGCTGCGGCT 3600
Db 3541 CAAACCTGAGGCTGCGGCTGAGCTGCGGCGCAGCTGCGGCTGCGGCTGCGGCTGCGGCT 3600
Qy 3601 CAGGCTCTGCTGGGCGAGCAATTCGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCA 3660
Db 3601 CAGGCTCTGCTGGGCGAGCAATTCGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCCA 3660
Qy 3661 GCGGCTGCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 3720
Db 3661 GCGGCTGCTGGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 3720
Qy 3721 GAAAGTGAAGCACTTCTGCTGCTGAAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 3780
Db 3721 GAAAGTGAAGCACTTCTGCTGCTGAAGGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 3780
Qy 3781 GGAATTCACCCGCTCAAGAGAGATGAGGCTCCAGAGCTGAGGCTGAGGCTGAGGCTGAG 3840
Db 3781 GGAATTCACCCGCTCAAGAGAGATGAGGCTCCAGAGCTGAGGCTGAGGCTGAGGCTGAG 3840
Qy 3841 GGGGCTCATGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACT 3900
Db 3841 GGGGCTCATGCTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACTTCTGACT 3900
Qy 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCCAACACTGTTGAGCCAGGAGACAGAGCTG 3960
Db 3901 CCCACTTTGTAGAGCCAGCCCTGTATGCCAACACTGTTGAGCCAGGAGACAGAGCTG 3960
Qy 3961 TGAGCTCTGGGCTTCTGAGCCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 4020
Db 3961 TGAGCTCTGGGCTTCTGAGCCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 4020
Qy 4021 CCCACTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCTGCTCTCTCTCTCTCTCT 4080
Db 4021 CCCACTCCCAAGGCTTACCGAGCTGGGAGGAGTACAGTAGGCTGCTCTCTCTCTCTCTCT 4080
Qy 4081 TTTCTACAGGAAGTCACTCTCGAGGAGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGT 4140
Db 4081 TTTCTACAGGAAGTCACTCTCGAGGAGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGT 4140
Qy 4141 CATGGCT 4200
Db 4141 CATGGCT 4200
Qy 4201 CATCTGAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 4260
Db 4201 CATCTGAGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 4260
Qy 4261 TTGCTGCCCCACT 4320
Db 4261 TTGCTGCCCCACT 4320
Qy 4321 TCCAAATTCAGGAAACAAACATGTTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 4380
Db 4321 TCCAAATTCAGGAAACAAACATGTTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 4380
Qy 4381 CAGGTTTGGGCTCTGCAATTTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 4440
Db 4381 CAGGTTTGGGCTCTGCAATTTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 4440

```
QY 4441 AGGGCAGGACAGACCCAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500
Db 4441 AGGGCAGGACAGACCCAGACCCAGAGTCTCCAAAGCACAGAGTGGCAACAAA 4500
QY 4501 ACCCAGCTGAGCATCAGACCTTGCTCGAATTGCTTCCAGTATTACGGTGCCTCTTC 4560
Db 4501 ACCCAGCTGAGCATCAGACCTTGCTCGAATTGCTTCCAGTATTACGGTGCCTCTTC 4560
QY 4561 TCTGCCCTCTTCCAGGATATCTGTGGGTGTCAGGCTGGGAGGCAACATAGCCAC 4620
Db 4561 TCTGCCCTCTTCCAGGATATCTGTGGGTGTCAGGCTGGGAGGCAACATAGCCAC 4620
QY 4621 ACCCAGGATTTCTGAAAGTTTAAATGAGTAGCATTTTGGGGTGTAGGTTGGCAGCT 4680
Db 4621 ACCCAGGATTTCTGAAAGTTTAAATGAGTAGCATTTTGGGGTGTAGGTTGGCAGCT 4680
QY 4681 CCCCAGGCTCTGCCCTCCAGCCCAACCCACTCATGACTCTAAAGTGTGTTGTTAATAT 4740
Db 4681 CCCCAGGCTCTGCCCTCCAGCCCAACCCACTCATGACTCTAAAGTGTGTTGTTAATAT 4740
QY 4741 TTATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTGCGAATTTCTATTCTTGT 4800
Db 4741 TTATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTGCGAATTTCTATTCTTGT 4800
QY 4801 TTAACAAATAAATGTTGCCCCAGAACTTAGTCTCTTTGCCAGGCTCACCCCTCTCTGG 4860
Db 4801 TTAACAAATAAATGTTGCCCCAGAACTTAGTCTCTTTGCCAGGCTCACCCCTCTCTGG 4860
QY 4861 TGCTCATGAGACTCTTGCCACCCCTTGCTCTCCACTCCCTGCTTGCCTCTGTTGGAGCTGC 4920
Db 4861 TGCTCATGAGACTCTTGCCACCCCTTGCTCTCCACTCCCTGCTTGCCTCTGTTGGAGCTGC 4920
QY 4921 ACAGAGCTCTGGAAAGAGGCTCTTCTCTCCCGACCTGGGGGCGATGGGGCGACCTCAGA 4980
Db 4921 ACAGAGCTCTGGAAAGAGGCTCTTCTCTCCCGACCTGGGGGCGATGGGGCGACCTCAGA 4980
QY 4981 CTTTACCACCTGCTGCTGCCACCAACCCCTTCTGATCCCTCAGTCTCTCCCAACAGCTTC 5040
Db 4981 CTTTACCACCTGCTGCTGCCACCAACCCCTTCTGATCCCTCAGTCTCTCCCAACAGCTTC 5040
QY 5041 TGTCCACCCAGGTTTCCCTCACCCACCTTTGCTAAAGTCTCTCTCA 5087
Db 5041 TGTCCACCCAGGTTTCCCTCACCCACCTTTGCTAAAGTCTCTCTCA 5087

RESULT 2
US-08-988-819-3
; Sequence 3, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED
; OTHER INFORMATION: REGION OF >1KB"
; US-08-988-819-3
```

```
Query Match 100.0%; Score 5087; DB 3; Length 5087;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCTGCGAGTGGGCTGGCGGATGGGATTATTAAAGCTTCGCCGAGCGCGGCTCGCC 60
Db 1 AACCTGCGAGTGGGCTGGCGGATGGGATTATTAAAGCTTCGCCGAGCGCGGCTCGCC 60
QY 61 CTCCCACTCCCGCAGCCTTCGGGAGAGGAGCGCACCCGCGCGGCCAGCCCCAGCCCCA 120
Db 61 CTCCCACTCCCGCAGCCTTCGGGAGAGGAGCGCACCCGCGCGGCCAGCCCCAGCCCCA 120
QY 121 TGGACCTCGAGAGGTTGAAAAACCAAACTAGCCCTCTCTTCAATACATGACAAGCAG 180
Db 121 TGGACCTCGAGAGGTTGAAAAACCAAACTAGCCCTCTCTTCAATACATGACAAGCAG 180
QY 181 CGCCCACTCTGATACCTAAACCGACCAAGTCAACAGCCCTTCCAACTCAGCCCTGTC 240
Db 181 CGCCCACTCTGATACCTAAACCGACCAAGTCAACAGCCCTTCCAACTCAGCCCTGTC 240
QY 241 CAGACCTCACCACATCTTGTGTGAGTCTCAAACTCAACCGCAGCTAAATCAACCAATCCCA 300
Db 241 CAGACCTCACCACATCTTGTGTGAGTCTCAAACTCAACCGCAGCTAAATCAACCAATCCCA 300
QY 301 AGTCTAAACTAATCTGAAACTTTTAAAGTAACCCAGTCTTAAACCTAAGTACCCCAAT 360
Db 301 AGTCTAAACTAATCTGAAACTTTTAAAGTAACCCAGTCTTAAACCTAAGTACCCCAAT 360
QY 361 GCCAATTATCTACCTAGCCAAACCCCTTGTCCAGTCCAAAGTGTCCACTG 420
Db 361 GCCAATTATCTACCTAGCCAAACCCCTTGTCCAGTCCAAAGTGTCCACTG 420
QY 421 AATCTCTACCTTGTGCTCTCACTGAAATCCAGAAAAGCATATTTCCCACTGCCACAT 480
Db 421 AATCTCTACCTTGTGCTCTCACTGAAATCCAGAAAAGCATATTTCCCACTGCCACAT 480
QY 481 CCCTCTTACAGCACCAACCCCTGCGCTCTGGACTCTCTGGTATCTCTGGGATGTCCAAACT 540
Db 481 CCCTCTTACAGCACCAACCCCTGCGCTCTGGACTCTCTGGTATCTCTGGGATGTCCAAACT 540
QY 541 CTGCAGTGCATCAGCCCAACAGCCGAGCTGTGTAATGACACCTCTCTCCCTCTGTCC 600
Db 541 CTGCAGTGCATCAGCCCAACAGCCGAGCTGTGTAATGACACCTCTCTCCCTCTGTCC 600
QY 601 CACCCCTTGCAGGCTGATGGAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAACACC 660
Db 601 CACCCCTTGCAGGCTGATGGAAGGCTCATTTGAAGTCCAACTTTTCCCACTTAACACC 660
QY 661 AAGAACGGGTGAACCTCCACACTGCCCACCGTTCCCTTGAGAGTGAGCAGTCAATCTCTT 720
Db 661 AAGAACGGGTGAACCTCCACACTGCCCACCGTTCCCTTGAGAGTGAGCAGTCAATCTCTT 720
QY 721 CAATCTAAACCCCACTTACCTTCCCACTCTAGGAATCAGTCTAGATATATACCCAAA 780
Db 721 CAATCTAAACCCCACTTACCTTCCCACTCTAGGAATCAGTCTAGATATATACCCAAA 780
```


Db 2941 CGTGTGCTTCTCGAACTCCTCTTCTTGCTCTCTGCACTCTCTCGGCGCCACCTTCTA 3000
Qy 3001 CCTCTCAGCTCCTCAGGCTCCTCTTCTCTCTCTGCTCTGCCCCCAGAGGGGCACTCTC 3060
Db 3001 CCTCTCAGCTCCTCAGGCTCCTCTTCTCTCTCTGCTCTGCCCCCAGAGGGGCACTCTC 3060
Qy 3061 CCAAGTTTGGCCCA CCGAGCCAAATCAGACGTCCTTCTCTGAGGCTCTTGCTGCGTCTCTC 3120
Db 3061 CCAAGTTTGGCCCA CCGAGCCAAATCAGACGTCCTTCTCTGAGGCTCTTGCTGCGTCTCTC 3120
Qy 3121 CTCTCTCTTTTCTACGCTCTCCAATGAGAGCTCACACCGCCCACTGCTTCAAATGTC 3180
Db 3121 CTCTCTCTTTTCTACGCTCTCCAATGAGAGCTCACACCGCCCACTGCTTCAAATGTC 3180
Qy 3181 ACCTGCATCAAAATGATATCTTATTTGAAAACCTCAGGGAGCCATGACAAAGAGCC 3240
Db 3181 ACCTGCATCAAAATGATATCTTATTTGAAAACCTCAGGGAGCCCAATGAAACAAAGAGCC 3240
Qy 3241 TAGCATGGAGACAGGCGCAGTGTGAGGGGACACAAAATAGAAAATTTGGGAGCAGGTA 3300
Db 3241 TAGCATGGAGACAGGCGCAGTGTGAGGGGACACAAAATAGAAAATTTGGGAGCAGGTA 3300
Qy 3301 TCTCTCTGCTGTGAGCCAGCGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360
Db 3301 TCTCTCTGCTGTGAGCCAGCGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360
Qy 3361 ACAGCTGAATACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGG 3420
Db 3361 ACAGCTGAATACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCTCTCCCGCTGGG 3420
Qy 3421 GGCAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480
Db 3421 GGCAGAGACTCTGCCAGGGCCACTGTGACTTGGAGGTGTGGCGAAGCCCTCAATGACAA 3480
Qy 3481 ACTGCGGCTGACCCAGAACTACAGGCTTACAGCCACCTTCTGTGTACTTGGTGGCCCT 3540
Db 3481 ACTGCGGCTGACCCAGAACTACAGGCTTACAGCCACCTTCTGTGTACTTGGTGGCCCT 3540
Qy 3541 CAACTGTAGGCTGACCTGCTGAGTGTGCGCGGAGCTTGGCGGCTTGGCGGCTTGGCGGCT 3600
Db 3541 CAACTGTAGGCTGACCTGCTGAGTGTGCGCGGAGCTTGGCGGCTTGGCGGCTTGGCGGCT 3600
Qy 3601 CAGGCGCTGCTGGGAGCAATTCGGGCGTCTATGGGAGCTCTGGGCTTACCCACTGCCCA 3660
Db 3601 CAGGCGCTGCTGGGAGCAATTCGGGCGTCTATGGGAGCTCTGGGCTTACCCACTGCCCA 3660
Qy 3661 GCGCTGCTGGGAGCTGAACCCACTTGGACTTGGAGCTTGGAGCTTGGAGCTTGGAGCTTGG 3720
Db 3661 GCGCTGCTGGGAGCTGAACCCACTTGGACTTGGAGCTTGGAGCTTGGAGCTTGGAGCTTGG 3720
Qy 3721 GAAGATGGAGCACTTGGGCTGTGAGGAGCTGACAGCTGCGGCTGCGGCTGCGGCTGCGGCT 3780
Db 3721 GAAGATGGAGCACTTGGGCTGTGAGGAGCTGACAGCTGCGGCTGCGGCTGCGGCTGCGGCT 3780
Qy 3781 GGAATTTCAACCGGCTCAAGAAAGAGTGCAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAG 3840
Db 3781 GGAATTTCAACCGGCTCAAGAAAGAGTGCAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAG 3840
Qy 3841 GGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900
Db 3841 GGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3900
Qy 3901 CCCACTTGTGAGAGCCAGCCCTGTATGCCAAACCTGTTGAGCCAGGAGCAGAGCTG 3960
Db 3901 CCCACTTGTGAGAGCCAGCCCTGTATGCCAAACCTGTTGAGCCAGGAGCAGAGCTG 3960
Qy 3961 TGAGCTCTGGCCCTTTCTGAGCCGCTGGCGCTGTATGCCATCAGCCCTGTCTCTCT 4020
Db 3961 TGAGCTCTGGCCCTTTCTGAGCCGCTGGCGCTGTATGCCATCAGCCCTGTCTCTCTCT 4020
Qy 4021 CCCACTTCCAAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCGCTGTCTGTCTCTG 4080
Db 4021 CCCACTTCCAAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCGCTGTCTGTCTCTG 4080

Qy 4081 TTTTCTACGAAAGTCACTGCTCGAGGAGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGT 4140
Db 4081 TTTTCTACGAAAGTCACTGCTCGAGGAGTGTGAAGTGTGTGAGTGTGTGAGTGTGTGAGTGTGT 4140
Qy 4141 CATGGCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 4200
Db 4141 CATGGCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 4200
Qy 4201 CATCTGAGGAGGAGGCTTGGGGGCGCCACACACACATGCTTCTTCTGCGGTGAAGCCCT 4260
Db 4201 CATCTGAGGAGGAGGCTTGGGGGCGCCACACACACATGCTTCTTCTGCGGTGAAGCCCT 4260
Qy 4261 TTGGCTGCCCACTCTCTCTTGGATGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4320
Db 4261 TTGGCTGCCCACTCTCTCTTGGATGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4320
Qy 4321 TCCAAATTCAGGAAACAAACATGCTGGGCAATTTCTACAAAAGAGATGAGATTACAGTG 4380
Db 4321 TCCAAATTCAGGAAACAAACATGCTGGGCAATTTCTACAAAAGAGATGAGATTACAGTG 4380
Qy 4381 CAGGTTGGGCTCTGCAATGGAGGTGCCCTATAAACCAAGAGAGAAATACCTGAAAGCAC 4440
Db 4381 CAGGTTGGGCTCTGCAATGGAGGTGCCCTATAAACCAAGAGAGAAATACCTGAAAGCAC 4440
Qy 4441 AGGGGCAAGGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 4500
Db 4441 AGGGGCAAGGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 4500
Qy 4501 ACCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGTCTTCCAGTATTAGGTTGCTTCT 4560
Db 4501 ACCGAGCTGAGCATCAGGACCTTGGCTCGAATTTGTCTTCCAGTATTAGGTTGCTTCT 4560
Qy 4561 TCTGCCCTTTTCCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4620
Db 4561 TCTGCCCTTTTCCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4620
Qy 4621 ACCACAGGATTTCTGAAAGTTTACAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 4680
Db 4621 ACCACAGGATTTCTGAAAGTTTACAATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 4680
Qy 4681 CCCCAAGGCTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740
Db 4681 CCCCAAGGCTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4740
Qy 4741 TTAATTTTTCGAGATGTTATTTATAGATGATATTTATTTAGATGATATTTATTTAGATGAT 4800
Db 4741 TTAATTTTTCGAGATGTTATTTATAGATGATATTTATTTAGATGATATTTATTTAGATGAT 4800
Qy 4801 TTAACAAATAAAATGCTTGGCCCGAGAACTTAGTCTCTTTTGGCCAGCCTCACCCCTCTCTGG 4860
Db 4801 TTAACAAATAAAATGCTTGGCCCGAGAACTTAGTCTCTTTTGGCCAGCCTCACCCCTCTCTGG 4860
Qy 4861 TGCTCATCAGACTCTTGGCCAGCCTTGGCTTCCACCTTCCCTGCTTGGCTTGGTGGAGCTGC 4920
Db 4861 TGCTCATCAGACTCTTGGCCAGCCTTGGCTTCCACCTTCCCTGCTTGGCTTGGTGGAGCTGC 4920
Qy 4921 ACAGAGCTCTGGGAGAGGCGCTTCTTCTTCCCGCAGCTGGGCGGATGGGCGCAGCTCAG 4980
Db 4921 ACAGAGCTCTGGGAGAGGCGCTTCTTCTTCCCGCAGCTGGGCGGATGGGCGCAGCTCAG 4980
Qy 4981 CTTACCCACTGCTGCTGCCACCAACCAACCCCTTGTGATTCCTCAGTCTCTCCACAGCTTTC 5040
Db 4981 CTTACCCACTGCTGCTGCCACCAACCAACCCCTTGTGATTCCTCAGTCTCTCCACAGCTTTC 5040
Qy 5041 TGTCCACCCAGGTTTCTTCCCTCACCCCACTTTTGTAAAGTCTTCTCTCA 5087
Db 5041 TGTCCACCCAGGTTTCTTCCCTCACCCCACTTTTGTAAAGTCTTCTCTCA 5087

RESULT 3

US-09-016-534-3

; Sequence 3, Application US/09016534

Db	2521	CTGCAACCCCTGCTCCTATACAGCTGAACCCCTTCTTTTCGGAGTGTTAGTAGTACCCCGTCTC	2580
Qy	2581	TCCCCAGCCCTCAGCTGGTGGCCCTGGGTGTGTGTCAGCGGCAAAATGGGGCTCTGGTTCCA	2640
Db	2581		2640
Db	2581	TCCCCAGCCCTCAGCTGGTGGCCCTGGGTGTGTGTCAGCGGCAAAATGGGGCTCTGGTTCCA	2640
Qy	2641	ATGGGCCACTCTCATCTCTCTCTGTTCTCTGTGAGAAAACTTTGCTTCACTCCACTG	2700
Db	2641		2700
Qy	2701	CCCTCTAGATTTCCGACCCCTTTTCTCTCTGGCTTCCCTGCCAAAATTTCTCCCAAGGA	2760
Db	2701		2760
Qy	2761	GTGGTCTACACCCTCTGCGCTCCACTTCCTCTCACCCCACTCACTTCTTAAACCCCTGCAA	2820
Db	2761		2820
Qy	2821	TCTGGCTTCAGGCCCCAGCAATGTTTCTCTCAAGGTGTGAGGACCTTCTTTGCCAAG	2880
Db	2821		2880
Qy	2881	CCCGACAGTGTTTGAAGGCTCAATTCCTCTGCTGTCTGTTTTCAGCCACACTGCTGAG	2940
Db	2881		2940
Qy	2941	CGCTGTGCTTCTCGAACTCCTCTTCTCTGGTCTCTGCACTCTCTCTGGGCCACTTCTA	3000
Db	2941		3000
Qy	3001	CCTCTCCAGTCTCTCAGGCTCTCTTCTCTCTGTCTGTGCTGCGCCACAGCGGGCACTCTC	3060
Db	3001		3060
Qy	3061	CGAAGTTTGCACCCAGCCCAATCAGCAGTCTCTCTCTGAGGTCTTGTGCGTCTCTCTC	3120
Db	3061		3120
Qy	3121	CTCCTCTTTTCTAAGCCCTCTCCATTTGGAGAGCTCAACCCGCCACTGTCTCAACTGTCT	3180
Db	3121		3180
Qy	3181	ACCTGCATCAAAATGATATCTTATTGGAAAACTCAGGGAGGCCATGAACAAAGAGCC	3240
Db	3181		3240
Qy	3241	TAGCATGGAGACAGGGCCAGTGTGTCAGGGGACACAAAAATAGAAAATTTGGAGCAGGTA	3300
Db	3241		3300
Qy	3301	TCTCTTGTGTGTGAGCAGCGGCTCTGCCCTCTCTCTCTCCCTACCCCTCTCTTTTC	3360
Db	3301		3360
Qy	3361	ACAGCTGAACTPACTGGGGCCCCCTTTCAAGCGCAGACTTCAACCTCTCCCGCTGGG	3420
Db	3361		3420
Qy	3421	GGCAGACTCTGCCACAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAA	3480
Db	3421		3480
Qy	3481	ACTCGGCTGACCCAGAACTACGAGGCCCTACAGCCACTTCTGTGTTACTTGGTGGCCT	3540
Db	3481		3540
Qy	3541	CAACCGTCAAGCGTGCACCTGTGAGCTGCGCGCAGCTGGCCCACTTCTGCACCAAGCCT	3600
Db	3541		3600
Qy	3601	CCAGGCCCTGTGGGACGATTTGGGGGGCTCATGCGAGCTCTGGGTCACTCCACTGCCCA	3660
Db	3601		3660

Qy	3661	GCCGCTGCTGGGACTGAACCCACTTTGGACTCTGGCCCTGCCCAAGTGAATTCCTCCA	3720
Db	3661		3720
Qy	3721	GCCGCTGCTGGGACTGAACCCACTTTGGACTCTGGCCCTGCCCAAGTGAATTCCTCCA	3780
Db	3721		3780
Qy	3781	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	3840
Db	3781		3840
Qy	3841	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	3900
Db	3841		3900
Qy	3901	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	3960
Db	3901		3960
Qy	3961	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4020
Db	3961		4020
Qy	4021	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4080
Db	4021		4080
Qy	4081	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4140
Db	4081		4140
Qy	4141	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4200
Db	4141		4200
Qy	4201	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4260
Db	4201		4260
Qy	4261	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4320
Db	4261		4320
Qy	4321	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4380
Db	4321		4380
Qy	4381	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4440
Db	4381		4440
Qy	4441	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4500
Db	4441		4500
Qy	4501	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4560
Db	4501		4560
Qy	4561	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4620
Db	4561		4620
Qy	4621	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4680
Db	4621		4680
Qy	4681	GAGATGGAAGACATCTTGGGCTGTGAAGAGAGCTGCAGACTGGCTGTGGCGCTGGGCCAA	4740
Db	4681		4740

RESULT 4

```

US-09-106-182-1
; Sequence 1, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 46..720
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 46..126

```

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 127..720
US-09-106-182-1

Query Match 28.8%; Score 1463.4; DB 3; Length 1710;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1464; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	3363	ACGTGAACCTACCTGGCCCTTTCACAGCCAGACTTCAACCCCTCCCGCTGGGG	3422
DB	227	ATCTGAACCTACCTGGCCCTTTCACAGCCAGACTTCAACCCCTCCCGCTGGGG	286
QY	3423	CAGAGACTCTGCCAGGCGCACTTGTGACTTGGAGGTGGCGAAGGCTCAATGACAAAC	3482
DB	287	CAGAGACTCTGCCAGGCGCACTTGTGACTTGGAGGTGGCGAAGGCTCAATGACAAAC	346
QY	3483	TGCGGTGACCCAGAACTACGAGCCCTACAGCCACCTTGTGTTACTTCGTGGCTCA	3542
DB	347	TGCGGTGACCCAGAACTACGAGCCCTACAGCCACCTTGTGTTACTTCGTGGCTCA	406
QY	3543	ACGTCAGGCTGCCACTGCTAGCTGGCGGCGAGCCTGGCCCACTTCTGCACAGCCCTCC	3602
DB	407	ACGTCAGGCTGCCACTGCTAGCTGGCGGCGAGCCTGGCCCACTTCTGCACAGCCCTCC	466
QY	3603	AGGCGCTGCTGGGCGAGCTTGGCGGCGTATGGCAGCTCTGGGCTACCCACTGCCCCAGC	3662
DB	467	AGGCGCTGCTGGGCGAGCTTGGCGGCGTATGGCAGCTCTGGGCTACCCACTGCCCCAGC	526
QY	3663	CGTGTCTGGGACTGAACCCACTTGGACTCTCGCCCTGCCACAGTGACTTCCTCAGA	3722
DB	527	CGTGTCTGGGACTGAACCCACTTGGACTCTCGCCCTGCCACAGTGACTTCCTCAGA	586
QY	3723	AGATGACGACTTCTGGCTGCTAGGAGCTGCAGACTGGCTGTGGGCTCGGCCAAGG	3782
DB	587	AGATGACGACTTCTGGCTGCTAGGAGCTGCAGACTGGCTGTGGGCTCGGCCAAGG	646
QY	3783	ACTTCAACCGGCTCAAGAAAGATGAGCTCCAGCAGCTGCACTGCACTCCCTGCACCTGG	3842
DB	647	ACTTCAACCGGCTCAAGAAAGATGAGCTCCAGCAGCTGCACTGCACTCCCTGCACCTGG	706
QY	3843	GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTTCTCGCTCCCTTCAAACCTGTCTCC	3902
DB	707	GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTTCTCGCTCCCTTCAAACCTGTCTCC	766
QY	3903	CACCTTGTGAGAGCCAGCCCTGTATGCCAACCTGTGTGAGCCAGGAGACAGAGCTGTG	3962
DB	767	CACCTTGTGAGAGCCAGCCCTGTATGCCAACCTGTGTGAGCCAGGAGACAGAGCTGTG	826
QY	3963	AGCCTCTGGCCCTTCTCTGACCGGCTGGGCTGTGATGCGATCAGCCCTGTCTCTCTCC	4022
DB	827	AGCCTCTGGCCCTTCTCTGACCGGCTGGGCTGTGATGCGATCAGCCCTGTCTCTCTCC	886
QY	4023	CACCTCCCAAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTGT	4082
DB	887	CACCTCCCAAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTCTGT	946
QY	4083	TCTACAGGAAGTCACTGCTCAGGAGGTGTGAGTGGTTCAGTGGTTCAGAGGGCTCA	4142
DB	947	TCTACAGGAAGTCACTGCTCAGGAGGTGTGAGTGGTTCAGTGGTTCAGAGGGCTCA	1006
QY	4143	TGGCTCTCTCTCTTCTGCTTACCTTGGCCAGTGGCCAGCCAGCCCTCAGGTGGCACA	4202
DB	1007	TGGCTCTCTCTCTTCTGCTTACCTTGGCCAGTGGCCAGCCAGCCCTCAGGTGGCACA	1066
QY	4203	TCTGGAGGCGAGGGGTTGAGGGGCAACACACATGCCCTTCTGGGGTGAAGCCCTTT	4262
DB	1067	TCTGGAGGCGAGGGGTTGAGGGGCAACACACATGCCCTTCTGGGGTGAAGCCCTTT	1126
QY	4263	GGCTGCCCACTCTCTCTGGATGGGTGTGCTCTCTTATCCCAATCACTCTATACATC	4322
DB	1127	GGCTGCCCACTCTCTCTGGATGGGTGTGCTCTCTTATCCCAATCACTCTATACATC	1186

QY	4323	CAATTAGGAAAACAAACATGGTGGCAATTTTACACAAAAGAGATGAGATTAACTGCA	4382
DB	1187	CAATTAGGAAAACAAACATGGTGGCAATTTTACACAAAAGAGATGAGATTAACTGCA	1246
QY	4383	GGGTTGGGTTCTGCAATTTGGAGGTGCCCTATAAACAGAAAGAGAAATACTGAAAGCACAG	4442
DB	1247	GGGTTGGGTTCTGCAATTTGGAGGTGCCCTATAAACAGAAAGAGAAATACTGAAAGCACAG	1306
QY	4443	GGGCGAGGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC	4502
DB	1307	GGGCGAGGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC	1366
QY	4503	CCGAGCTGAGCATCAGGACCTTGCCTCGAATGTCTCCAGTATTAACGTCCTCTTCTC	4562
DB	1367	CCGAGCTGAGCATCAGGACCTTGCCTCGAATGTCTCCAGTATTAACGTCCTCTTCTC	1426
QY	4563	TGCCCTCTTCCAGGGTATCTGTGGGTGCGAGCTCGGGAGGCAACCATAGCCACAC	4622
DB	1427	TGCCCTCTTCCAGGGTATCTGTGGGTGCGAGCTCGGGAGGCAACCATAGCCACAC	1486
QY	4623	CACAGGATTTCTGAAAGTTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA	4682
DB	1487	CACAGGATTTCTGAAAGTTTACAAATGCAATGCAATGCAATGCAATGCAATGCAATGCA	1546
QY	4683	CCAAAGGCTGCTGCCCTCCAGCCCACTCATGACTCTAAGTGTGTGTTAATTAATTT	4742
DB	1547	CCAAAGGCTGCTGCCCTCCAGCCCACTCATGACTCTAAGTGTGTGTTAATTAATTT	1606
QY	4743	ATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTAGCAATTTCTTCTCTGATT	4802
DB	1607	ATTTATTTGGAGATGTTATTTATTTAGATGATATTTATTTAGCAATTTCTTCTCTGATT	1666
QY	4803	AACAAATAAAATGCTTGCCCCAGAA	4827
DB	1667	AACAAATAAAATGCTTGCCCCAGAA	1691

RESULT 5

US-08-792-019B-1

; Sequence 1, Application US/08792019B

; Patent No. 5741772

; GENERAL INFORMATION:

; APPLICANT: CHANG, MING-SHI

; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMGEN INC.

; STREET: 1840 DEHAVILLAND DRIVE

; CITY: THOUSAND OAKS

; STATE: CA

; COUNTRY: USA

; ZIP: 91320

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/792,019B

; FILING DATE: 03-FEB-1997

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: COOK, ROBERT R.

; REGISTRATION NUMBER: 31,602

; REFERENCE/DOCKET NUMBER: A-442

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 797 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; FEATURE:

APPLICANT: CHANG, MING-SHI
APPLICANT: ELLIOTT, GARY S.
APPLICANT: SARMIENTO, ULLA
APPLICANT: SENALDI, GIORGIO
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC.
STREET: ONE AMGEN CENTER
CITY: THOUSAND OAKS
STATE: CA
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,534
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/792,019
FILING DATE: 03-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: COOK, ROBERT R.
REGISTRATION NUMBER: 31,602
REFERENCE/DOCKET NUMBER: A-442B
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 90..764
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 171..764
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 90..170
US-09-016-534-1

Query Match 10.3%; Score 523.4; DB 3; Length 797;
Best Local Similarity 99.8%; Pred. No. 1.1e-120;
Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 3422
DB 271 ATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 330
QY 3423 CAGAGACTCTGCCAGGCCACTGTGTGACTTGGAGGTGGCGAGCCCTCAATGACAAAC 3482
DB 331 CAGAGACTCTGCCAGGCCACTGTGTGACTTGGAGGTGGCGAGCCCTCAATGACAAAC 390
QY 3483 TGGCGTGAACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTTACTTGGCGTGCCTCA 3542
DB 391 TGGCGTGAACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTTACTTGGCGTGCCTCA 450
QY 3543 ACCGTGAGGCTGCACTCTGAGCTGCGCGGAGCTGCGCCACTTCTGACACAGCCTCC 3602
DB 451 ACCGTGAGGCTGCACTCTGAGCTGCGCGGAGCTGCGCCACTTCTGACACAGCCTCC 510
QY 3603 AGGCGCTGCTGGGAGCAATTCGGCGCTGATGGCAAGCTCTGGGCTACCCACTGCCCCAGC 3662
DB 511 AGGCGCTGCTGGGAGCAATTCGGCGCTGATGGCAAGCTCTGGGCTACCCACTGCCCCAGC 570
QY 3663 CGTGCCTGGGACTGAACCCCACTGCACTCTGCGGCCCTGCCACAGTGAATCTCTCCAGA 3722

RESULT 8

US-08-792-019B-4
Sequence 4, Application US/08792019B
Patent No. 5741772
GENERAL INFORMATION:

APPLICANT: CHANG, MING-SHI
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC.
STREET: 1840 DEHAVILLAND DRIVE
CITY: THOUSAND OAKS
STATE: CA
COUNTRY: USA
ZIP: 91320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/792,019B
FILING DATE: 03-FEB-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: COOK, ROBERT R.
REGISTRATION NUMBER: 31,602
REFERENCE/DOCKET NUMBER: A-442

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 819 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 95..769
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 176..769
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 95..175
US-08-792-019B-4

Query Match 8.4%; Score 429.6; DB 1; Length 819;
Best Local Similarity 90.4%; Pred. No. 2.3e-97;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 3422
DB 276 ACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 335
QY 3423 CAGAGACTCTGCCAGGCCACTGTGTGACTTGGAGGTGGCGAGCCCTCAATGACAAAC 3482
DB 336 CAGAAACTCTGCCAGGCCACTGTGTGAGGTGGCGAGCCCTCAATGACAGGC 395

QY 3483 TCGGCTGACCCAGAACCTAGAGGCTCAGCCACCTTCTGTGTTACTTGGCGCTCA 3542
Db |||||
QY 396 TCGGCTGACCCAGAACCTATGAGGCTACAGTCACTCTCTGTGTTACTTGGCGCTCA 455
Db |||||
QY 3543 ACCGTGAGCTGCCACTGCTGAGCTGCGCGGAGCTGGCCACTTCTGTCAGCAGCTCC 3602
Db |||||
QY 456 ACCGTGAGCTGCCACTGCTGAGCTGCGCGGAGCTGGCCACTTCTGTCAGCAGCTCC 515
Db |||||
QY 3603 AGGGCTGCTGGGAGCACTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 3662
Db |||||
QY 516 AGGGCTGCTGGGAGCACTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 575
Db |||||
QY 3663 CGCTGCTGGGAGCACTTGGGGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 3722
Db |||||
QY 576 CTCTGCGAGGAGCTGAGCCAGCTTGGGGCTTGGGGCTTGGGGCTTGGGGCTTGGGGCT 635
Db |||||
QY 3723 AGATGAGCACTTCTGGCTGCTGAAGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 3782
Db |||||
QY 636 AGATGAGCACTTCTGGCTGCTGAAGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 695
Db |||||
QY 3783 ACTTCAACCGGCTCAAGAGAGAGTGCAGCCTCCAGCAGCTGCAGTCACTTGCACCTGG 3842
Db |||||
QY 696 ACTTCAACCGGCTCAAGAGAGAGTGCAGCCTCCAGCAGCTGCAGTCACTTGCACCTGG 755
Db |||||
QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870
Db |||||
QY 756 AGGCACATGGTTTCTGACCTCTGACCTT 783
Db |||||

RESULT 9

US-09-931-704-3
; Sequence 4, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 176..769

; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..175
US-08-988-819-4

Query Match 8.4%; Score 429.6; DB 3; Length 819;
Best Local Similarity 90.4%; Pred. No. 2.3e-97;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 3363 AGCTGAACCTACTCTGGGCCCCCTTTCAAGCAGCCAGACTTCAACCTCTCCCGCTGGGG 3422
Db |||||
QY 276 ACCTGAACCTACTCTGGGCCCCCTTTCAAGCAGCCAGACTTCAATCTCTCGACTGGGG 335
Db |||||
QY 3423 CAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGTGTGGAGAGCTCAATGACAAAC 3482
Db |||||
QY 336 CAGAACTCTGCCAGGGCCAGGTTCACTTGGAGTGTGGAGAGCTCAATGACAGGC 395
Db |||||
QY 3483 TCGGCTGACCCAGAACCTAGAGGCTTACAGCCACTTCTGTGTTACTTGGCGCTCA 3542
Db |||||
QY 396 TCGGCTGACCCAGAACCTATGAGGCTTACAGTCACTCTCTGTGTTACTTGGCGCTCA 455
Db |||||
QY 3543 ACCGTGAGCTGCCACTGCTGAGCTGCGCGGAGCTTGGCCACTTCTGACAGCCTCC 3602
Db |||||
QY 456 ACCGTGAGCTGCCACTGCTGAGCTTCCAGCTGAGCTTGGCCACTTCTGACAGCCTCC 515
Db |||||
QY 3603 AGGGCTGCTGGGAGCACTTGGCGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 3662
Db |||||
QY 516 AGGGCTGCTGGGAGCACTTGGCGGCTCATGGCAGCTTGGGCTACCCACTGCCCCAGC 575
Db |||||
QY 3663 CGTGGCTGGGAGCACTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 3722
Db |||||
QY 576 CTCTGCGAGGAGCTGAGCCAGCTTGGGGCTTGGGGCTTGGGGCTTGGGGCTTGGGGCT 635
Db |||||
QY 3723 AGATGAGCACTTCTGGCTGCTGAAGGAGCTGCAGCCTGCAGCAGCTTGGCTGGCGCT 3782
Db |||||
QY 636 AGATGAGCACTTCTGGCTGCTGAAGGAGCTGCAGCCTGCAGCAGCTTGGCTGGCGCT 695
Db |||||
QY 3783 ACTTCAACCGGCTCAAGAGAGAGTGCAGCCTCCAGCAGCTGCAGTCACTTGCACCTGG 3842
Db |||||
QY 696 ACTTCAACCGGCTTGAAGAGAGATGCAGCCTCCAGCAGCTGCAGTCACTTGCACCTGG 3870
Db |||||
QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870
Db |||||
QY 756 AGGCACATGGTTTCTGACCTCTGACCTT 783
Db |||||

RESULT 10

US-09-016-534-4
; Sequence 4, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 819 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 95..769

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 176..769

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 95..175

US-09-016-534-4

Query Match 8.4%; Score 429.6; DB 3; Length 819;
Best Local Similarity 90.4%; Pred. No. 2.3e-87; Indels 0; Gaps 0;
Matches 459; Conservative 0; Mismatches 49;

QY 3363 AGCTGTAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 3422

DB 276 ACCTGTAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 335

QY 3423 CAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGGAGGCTCAATGACAAAC 3482

DB 336 CAGAAACTCTGCCAGGCGCACTGTTGAAAGTGTGGGAGGCTCAATGACAGGC 395

QY 3483 TGGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTACTTGGTGGCTCA 3542

DB 396 TGGGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTACTTGGTGGCTCA 455

QY 3543 ACCGTGAGGCTGACCTGCTGAGTGGCGCGCAGCCCTGGCCACTTCTGCAACGACCTCC 3602

DB 456 ACCGTGAGGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 515

QY 3603 AGGCGCTGCTGGGAGCAATTCGGGCGCTCATGGAGCTCTGGGCTACCACTGCCCCAGC 3662

DB 516 AGGCGCTGCTGGGAGCAATTCGGGCGCTCATGGGAGCTGACCTGACCTGACCTGACCT 575

QY 3663 CGTGTGCTGGGAGCTGACCCCACTTGGGACTCTGCGCCCTGCGCCCACTGACCTTCTCCAGA 3722

DB 576 CTCTGCGAGGAGCTGACCCCACTTGGGACTCTGCGCCCTGCGCCCACTGACCTTCTCCAGA 635

QY 3723 AGATGAGCACTTCTGGCTGCTGAGGAGCTGACAGCTGCTGGCTGCTGGGCTGCGGCAAGG 3782

DB 636 AGATGAGCACTTCTGGCTGCTGAGGAGCTGACAGCTGCTGGCTGCTGGGCTGCGGCAAGG 695

QY 3783 ACTTCAACCGGCTCAAGAGAGATGAGCTTCCAGGAGCTGACCTGACCTGACCTGACCTGG 3842

DB 696 ACTTCAACCGGCTTAAAGAGAGATGAGCTTCCAGGAGCTGACCTGACCTGACCTGACCT 755

QY 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

DB 756 AGGCACATGGTCTTCTGACCTCTGACCT 783

RESULT 11

US-09-106-182-7

; Sequence 7; Application US/09106182

; Patent No. 6046035

; GENERAL INFORMATION:

; APPLICANT: Shi, Yanguu

; APPLICANT: Ruben, Steve

TITLE OF INVENTION: Cardiostrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PP385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-106-182-7

Query Match 6.4%; Score 324.6; DB 3; Length 396;

Best Local Similarity 95.2%; Pred. No. 2.1e-71;

Matches 374; Conservative 0; Mismatches 15; Indels 4; Gaps 4;

QY 3803 AAGATGAGCCCTCCAGCAGCTGCAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTT 3862

DB 4 ACGAGGAGCCCTCCAGCAGCTGCAGTCACTCCCTGACCTGGGGCTCATGGCTTCTGACTT 63

QY 3863 CTGACCTTCT 3922

DB 64 CTGACCTTCT 123

QY 3923 TGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTCTCTGG 3982

DB 124 TGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTCTCTGG 183

QY 3983 ACCGCTGGGCGTGTGATCGGATCAGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4042

DB 184 ACCGCTGGGCGTGTGATCGGATCAGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243

QY 4043 AGCTGGGAGGAGGTACAGTAGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4102

DB 244 AGCTGGGAGGAGGTACAGTAGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 303

QY 4103 AG-GGAGTGTGAAGTGG-TTCAGGTTGGTGCAGAGCGCTCATGGGCTCC-TGCTTCTTG 4159

DB 304 AGNGGAGTGTGAAGTGGTTCAGGTTGGTGCAGAGCGCTCATGGGCTCC-TGCTTCTTG 363

QY 4160 CCTACACTT-GGCCAGTGCACCCAGGCCCT 4191

DB 364 GCTACCANTGGGNCATGNCACACGACCTT 396

RESULT 12

US-08-232-463-14

; Sequence 14; Application US/08232463

```

: Patent No. 5670367
:
: GENERAL INFORMATION:
:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
:
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
:
: US-08-232-463-14

```

Qy	3023	CTCTTCCTCTCTGTCTGCCCCACAGCGGGCACTCTCCAAAGTTTGC	3071
		::: ::::: : : : : : : : : : : : : : : :	
Db	1400	YYYYYYYYYYYYYYYYYYYYYYYYYYYVYGGTACCAAATCTCTC	1448

```

RESULT 13
US-09-249-585A-4
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/00905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

```

Query Match	1.2%;	Score 60.2;	DB 4;	Length 1926;
Best Local Similarity	46.6%;	Pred. No. 2.4e-05;		
Matches 335;	Conservative 0;	Mismatches 373;	Indels 11;	Gaps

QY	2461	ATGCTGCTTTCCATAGACCTTCTCTCTGTCCTACTCATAGACGCTCCATTTCTTCTCTT	2520
DB	41	ATCTCTCTTCCCTCTGTGTAGACCTGGTCTTCCGAGGCGCGGTCACTCGAGTTTCTT	100
QY	2521	CTGCAACCTGTGCTTATCAGCTGAACCCCTTTTTCGGAGTGTTAGTAGAGTACCCGTC--	2578
DB	101	CTCCCCACTATTGGTACCTGCTCTCTGCCCCCTTCTCTGCTCTCTCCGCTTCTG	160
QY	2579	-TCFCCCCAGCCCTCAGCTGTGTGGGCTGGGGTGTGTAGCGGCAATATGGGGCTCTGGTT	2637
DB	161	GTCTCTGGGGCCCGCAGTCTAGTCCCGGTCTGTATCTATCCACAGGCCCTCTGGGG	220
QY	2638	CCAAATGGGCCACTCTCATCTCTCTGTGTCTCTGTGTCAGAAAACTTTGTCTTCACTCCA	2697
DB	221	TTTTTGCAGGTTCAACGTAACCGAGTTTCCCTGGGTGCCACCTTGCTCTGCTCGTC	280
QY	2698	CTGCCTCTCTAGTTCCGAGACCTTTTCTCTCTGGGCTTTCCCTGCGCAATTTCTCGAA	2757
DB	281	CTGCGCCTCCCGTCTCTGTCTCTCCCGGTCTCTGTCTCTCTCCCGTCTCTGCTCTCTC	340
QY	2758	GGAGTGGTCTTACACCTCT---CTGCTCTCACTTCTCTCTCCACCACTCACTTCTTAACTCC	2814
DB	341	CCCGTCTCTCCCGTCTCTCCCGTCTCTCCCGTCTCTCTCTCTCTCCCGTCTCTGCTCTC	400
QY	2815	CTGCAATCTGGCTTTCAGGCCCCAGAAATGTTCTCTCCAAAGTGTGTGAGGCACTCTCTT	2874
DB	401	CTCCCGTCTCTCCCGTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	460
QY	2875	GCCAAAGCCGACAGTGTTTTGAAGGCTCATTTCTCTTGTCTGTCTCTCTTTTTCAGCCACACT	2934
DB	461	GTCTCTCCCGTCTCTGTC	520
QY	2935	GCTGAGGCTGTGCTCTTCTGAACTCTCTCTTCTTGGTCTCTGTCACTCTCTCTGGGCGAC	2994
DB	521	CCCGTCTCTGTCTCTCTCCCGTCTCTCCCGTCTCTGTCTCTCTCTCTCTCTCTCTCTCTC	580
QY	2995	CTTCTACCTCTCCAGTCTCTCCAGTCTCTTCTCTCTGTCTGTGCTGCTGCCCAAGCGGGC	3054
DB	581	CTCCCGTCTCTGTC	639
QY	3055	ACTC-----TCCCAAGGTTTGCCCAAGCCCAATCAGCAGTCTTCTCTGAGGCTCTTGT	3110

[illegible]

RESULT 14

```

US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cell
; TITLE OF INVENTION: From Multiple
; FILE REFERENCE: 0867/1D903U1
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Vers
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

Query Match	1.2%	Score 60.2	DB 2	Length 1931	
Best Local Similarity	46.6%	Prod. No. 2.4e-05			
Matches 335	Conservative	0	Mismatches 373	Indels 11	Gaps
Qy	2461	ATGCTGTCTTCATAGACGCTTCCTCTGTCCTACTCATGAGACTGCTGCCTCATTTCTTCTCTT	2520		
Db	41	ATCTCTCTCTCTCTCTGTGTAGACCTGTGCTTCCGAGGCGCGGTACCTGTGAGTTCCTT	100		
Qy	2521	CTGCAACCTGTCTCTATACGCTGAACCTTCTTTTCGGAGTGTAGTGAGTACCCGTC--	2578		
Db	101	CTCCCCCACTATTGGTACCTGCTCTGCCCCCTTCTCTGCTCTGCTCTCTCGCCCTTCG	160		
Qy	2579	-TCTCCCAAGCCCTCAGCTGTGTGGCCTGGGTGTGACGGCAAAATGGGGCTCTGTGTT	2637		
Db	161	GTCTCGGGCGCGCGAGTCTAGTCCCGGTCTGTATCTCTACCCACAGGCTCTGTGGG	220		
Qy	2638	CCAAATGGGGCACTCTCATCTCTCTTGTTCCTTGTGACAGAAA	2697		
Db	221	TTTTTTCAGTTCAACGTAAACGACGTTTCCTTGGTGGCCACTTGTCTCTCTCTCGTC	280		
Qy	2698	CTGCCCTCTCTATGTTCCCGACCGCTTTTCTCTCTGGCTTTCTCTGCCAAAATTTCTCCAA	2757		
Db	281	CTCGCCCTCCCGTCTCTGCTCTCCCGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT	340		
Qy	2758	GGAGTGGTCTACACCT--CTGCCCTCACTTCTCTCCACCCACACACTCTCTTAACCCC	2814		
Db	341	CCGCTCTCTCCCGTCTCTCCCGTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCT	400		
Qy	2815	CTGCAATCTGGCTTCAGGCCCGCAGCAATGGTTCCTCTCCAAAGTCTCTCAGGCACCTCTT	2874		
Db	401	CTCCCGTCTCTCCCGCTCTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	460		
Qy	2875	GCCAAAGCCGACAGTGTTTTGAAGGGTCAATTCCTCTGTGTCTGTCTTTTGACGCCACACT	2934		
Db	461	GTCTCTCCCGTCT	520		
Qy	2935	CGTAGGGCTGTGCTTCTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2994		
Db	521	CCCGTCT	580		
Qy	2995	CTTCTACTCTCCAGCTCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	3054		
Db	581	CTTCCCGTCT	639		

[illegible]

RESULT 15

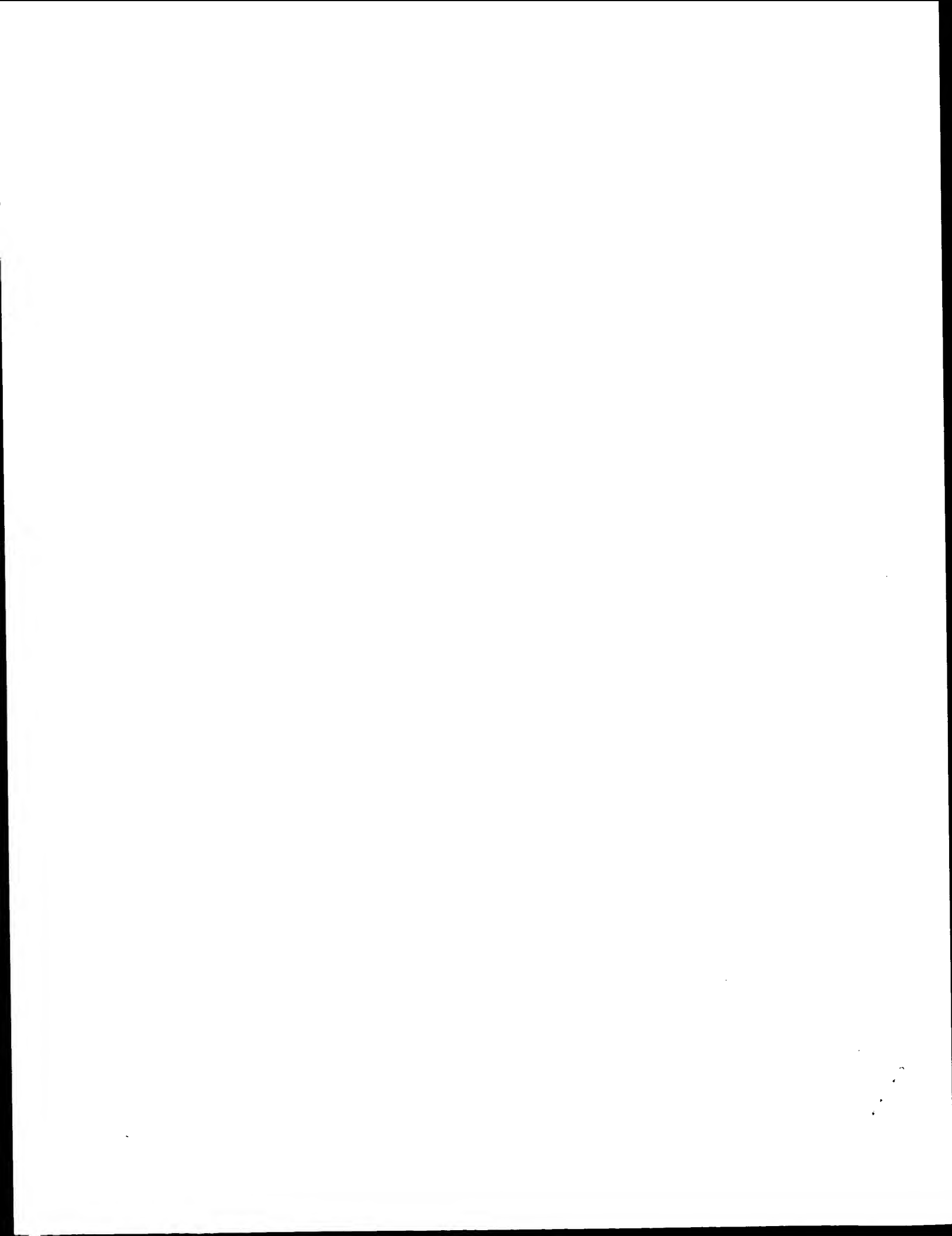
```

US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5326 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-658-136-2

```

	Query Match	1.1%;	Score 57.6;	DB 3;	Length 53526;
	Best Local Similarity	45.4%;	Pred. No. 0.00047;		
	Matches 367;	Conservative	0;	Mismatches 429;	Indels 12; Gaps
Qy	2319	CCTCCCGCTGGCGCCTTGCCCTATCCAGAGCCTCTCCCGTTGCTTCCCTTCTTCCAG	2378		
Db	34293	CCTCTCCCGCTCCCGCTAGCGCTTTCCCTTCTTCCCGCAGCGCTTCCCTCTCTCCGCTC	34352		
Qy	2379	TTATACATCTCGCTCATCCCTTTCCTGGGCGCCAGCGCTCCCGAGGGTTGGAAAGG	2438		
Db	34353	CCCTAGCGCTTTCCTCTCTCCCGCTCCCTACCGCTTCCCTCTCTCC--CCTCCCGCTAGAC	34410		
Qy	2439	GCTCTGGCCTCTTCCCTATACCATGCTGTCTCCATAGCGCTTCTCTGCTCCTACTATG	2498		
Db	34411	CTTTCGCTCACCTCTCCCGCTGAGCGCCCTCCACTGCTGCTCCCGAGCGCCCTCCCTCCCGCTA	34470		
Qy	2499	AGAGTGCCTCCATTTCTTCTCTGTGCAACCCCTGCTCTATCATAGCTGAAACCTCTTCTTCGG	2558		

Search completed: February 1, 2003, 08:50:00
Job time : 394.925 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 2017.75 Seconds
(without alignments)
11812.771 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 819
Sequence: 1 tattattaaagcttcgocgg.....aggccacgtcagctgtgctt 819

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inrv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.ly.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	819	100.0	819	6	AR002597	Sequence
2	819	100.0	819	6	AX392089	Sequence
3	819	100.0	819	10	AF176913	Mus muscu
4	674	82.3	881	6	AX205024	Sequence
5	674	82.3	881	6	AX205042	Sequence
6	669.4	81.7	797	6	AR002595	Sequence
7	669.4	81.7	797	6	AX392086	Sequence
8	669.4	81.7	797	9	AF176911	Homo sapi
9	654	79.9	1736	9	BC012939	Homo sapi
10	627	76.6	1689	9	AF172854	Homo sapi
11	590.4	72.1	680	9	AY049779	Homo sapi
12	540.8	66.0	269355	2	AC109138	Mus muscu
13	509.2	62.2	1692	6	AX205060	Sequence
14	429.6	52.5	5087	6	AR002596	Sequence
15	429.6	52.5	5087	6	AX392088	Sequence
16	429.6	52.5	5087	9	AF176912	Homo sapi
17	429.6	52.5	135116	2	AP002437	Homo sapi
18	429.6	52.5	168567	9	AP003419	Homo sapi
19	428	52.3	169144	2	AC005849	Homo sapi
20	236	28.8	283	6	AX202145	Sequence
21	137.6	16.8	63347	2	AC110526	Mus muscu
22	50.8	6.2	125020	9	AF429315	Homo sapi
23	46.2	5.6	47958	2	AC091104	Homo sapi
24	45.4	5.5	32134	2	AC096823	Rattus no
25	45.2	5.5	125020	9	AF429315	Homo sapi
26	44.4	5.4	187727	2	AC021142	Homo sapi
27	43.2	5.3	6803	9	HSU86758	Human netri
28	42.4	5.2	1357	9	AF035771	Homo sapi
29	42.4	5.2	1524	9	AF004900	Homo sapi
30	42.4	5.2	1578	6	AX410794	Sequence
31	42.4	5.2	1578	9	HSU82108	Human SIP-1
32	42.4	5.2	1600	9	HSTKALMR	ZS0150 H.sapiens m
33	42.4	5.2	1639	9	BC014513	Homo sapi
34	42.4	5.2	1642	6	AR070449	Sequence
35	42.2	5.2	1901	5	GGLAMB1	XL6878 Chicken mRN
36	42	5.1	63082	2	AC022663	Homo sapi
37	41.2	5.0	115666	2	AC105744	Oryza sat
38	41.2	5.0	119495	2	AC126302	Rattus no
39	41.2	5.0	123071	2	AC121481	Rattus no
40	41.2	5.0	173322	2	AC127095	Rattus no
41	41	5.0	158537	2	AC126887	Rattus no
42	40.8	5.0	1983	9	HSM803625	AL832318 Homo sapi
43	40.8	5.0	2750	9	AK097089	Homo sapi
44	40.8	5.0	4015	9	AB075829	Homo sapi
45	40.8	5.0	71418	9	AL732442	Human DNA

ALIGNMENTS

RESULT 1
AR002597
LOCUS
DEFINITION Sequence 4 from patent US 5741772.
ACCESSION AR002597
VERSION AR002597.1 GI:3964151
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 819)
AUTHORS Chang, M.-s.
TITLE Neurotrophic factor NNT-1
JOURNAL Patent: US 5741772-A 4 21-APR-1998;
FEATURES Location/Qualifiers

819 bp
DNA
linear PAT 04-DEC-1998


```
source 1. .819
BASE COUNT 156 a 288 c 218 g 157 t
ORIGIN
Query Match 100.0%; Score 819; DB 6; Length 819;
Best Local Similarity 100.0%; Pred. No. 3.4e-176;
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATTATTAAAGCTTCCCGGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCCCGGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGGAGA 60

Qy 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120

Qy 121 GGGGATGTTAGCTTATGCACTGCGGTGCTGTGGCACTCCCTCGAGTGCAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTATGCACTGCGGTGCTGTGGCACTCCCTCGAGTGCAGCTCTTAA 180

Qy 181 TCGACAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 TCGACAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

Qy 241 CCTGGAGCATCAACTCGGAGCTTAGCTGGGAGCTTACCTGAACTACCTGGGCGCCCTT 300
Db 241 CCTGGAGCATCAACTCGGAGCTTAGCTGGGAGCTTACCTGAACTACCTGGGCGCCCTT 300

Qy 301 CAACGAGCTGACTCAATCCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCACGGT 360
Db 301 CAACGAGCTGACTCAATCCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCACGGT 360

Qy 361 CAACCTTGGAAAGTGGGGAAGCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGC 420
Db 361 CAACCTTGGAAAGTGGGGAAGCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGC 420

Qy 421 GTACAGTCACTCTGTGTGTTACTTGGCTGAGCTCAACGCTCAGGCTGCCAGCTGAAC 480
Db 421 GTACAGTCACTCTGTGTGTTACTTGGCTGAGCTCAACGCTCAGGCTGCCAGCTGAAC 480

Qy 481 CCGAGCTAGCTGGCCCACTCTGTATACAGCTCCAGGGCGCTGTGGGAGCATTTGACG 540
Db 481 CCGAGCTAGCTGGCCCACTCTGTATACAGCTCCAGGGCGCTGTGGGAGCATTTGACG 540

Qy 541 TGTATGGCGACGCTTGGCTACCCACTGCCCCAGCCTGTGCGGAGCTGAGCCAGCCTG 600
Db 541 TGTATGGCGACGCTTGGCTACCCACTGCCCCAGCCTGTGCGGAGCTGAGCCAGCCTG 600

Qy 601 GGGCCCTGGCGCTGCCACAGTACTTCTCCAGAGATGATGACTTCTGGCTGCTGAA 660
Db 601 GGGCCCTGGCGCTGCCACAGTACTTCTCCAGAGATGATGACTTCTGGCTGCTGAA 660

Qy 661 GGAGCTGCAGACTGGCTGTATGGCTTACGCAAGGACTTCAACCGGCTTAAGAGAGAT 720
Db 661 GGAGCTGCAGACTGGCTGTATGGCTTACGCAAGGACTTCAACCGGCTTAAGAGAGAT 720

Qy 721 GCAGCTCCAGCAGCTTCACTGACCCCTGCACTTGGAGCAGATGTTTCTGACCTCTGAC 780
Db 721 GCAGCTCCAGCAGCTTCACTGACCCCTGCACTTGGAGCAGATGTTTCTGACCTCTGAC 780

Qy 781 CCTTAACCCCCACACCTCCAGGCGCCAGTCACTGTGCTT 819
Db 781 CCTTAACCCCCACACCTCCAGGCGCCAGTCACTGTGCTT 819

RESULT 2
AX392089 LOCUS AX392089 819 bp DNA linear PAT 23-MAR-2002
DEFINITION Sequence 4 from Patent WO0215977.
ACCESSION AX392089
VERSION AX392089.1 GI:19700577
KEYWORDS
```

```
Mus sp.
Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Senaldi, G.
TITLE Methods and compositions for treating ige-related disease using nn
JOURNAL t-1 inhibitors
Patent: WO 0215977-A 4 28-FEB-2002;
Amgen Inc. (US)

FEATURES
source Location/Qualifiers
1. .819
/organism="Mus sp."
/db_xref="taxon:10095"
95. .772
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD28986.1"
/db_xref="GI:19700578"
/translation="MDLRAGDSWGLACLTVMHLPAVPALNRTGDPGSPSIOKTY
DLTRYLEHQLRSLAGTYLNLGPPFPNPFPRIGATLPRATVNLVWRLNDRLLR
LTQNYEAYSHLLCYLRGLNQATAEELRSLAHFCTSLQGLIGSLAGMATLGYPLPQ
PLQETEPAWAPGPAHSDFLQKMDDFLLKELQTLWRSKDFNRLKKKQPPAASVTL
HLEAHGF"
sig_peptide 95. .175
mat_peptide 176. .769
/product="unnamed"

BASE COUNT 156 a 288 c 218 g 157 t
ORIGIN

Query Match 100.0%; Score 819; DB 6; Length 819;
Best Local Similarity 100.0%; Pred. No. 3.4e-176;
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATTATTAAAGCTTCCCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCCCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGA 60

Qy 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120

Qy 121 GGGGATGTTAGCTTATGCACTGCGGTGCTGTGGCACTCCCTCGAGTGCAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTATGCACTGCGGTGCTGTGGCACTCCCTCGAGTGCAGCTCTTAA 180

Qy 181 TCGACAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 TCGACAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

Qy 241 CCTGGAGCATCAACTCGGAGCTTAGCTGGGAGCTTACCTGAACTACCTGGGCGCCCTT 300
Db 241 CCTGGAGCATCAACTCGGAGCTTAGCTGGGAGCTTACCTGAACTACCTGGGCGCCCTT 300

Qy 301 CAACGAGCTGACTCAATCCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCACGGT 360
Db 301 CAACGAGCTGACTCAATCCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCACGGT 360

Qy 361 CAACCTTGGAAAGTGGGGAAGCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGC 420
Db 361 CAACCTTGGAAAGTGGGGAAGCTCAATGACAGGCTCGGCTGACCCAGAACTATGAGC 420

Qy 421 GTACAGTCACTCTGTGTGTTACTTGGCTGAGCTCAACGCTCAGGCTGCCAGCTGAAC 480
Db 421 GTACAGTCACTCTGTGTGTTACTTGGCTGAGCTCAACGCTCAGGCTGCCAGCTGAAC 480

Qy 481 CCGAGCTAGCTGGCCCACTCTGTATACAGCTCCAGGGCGCTGTGGGAGCATTTGACG 540
Db 481 CCGAGCTAGCTGGCCCACTCTGTATACAGCTCCAGGGCGCTGTGGGAGCATTTGACG 540

Qy 541 TGTATGGCGACGCTTGGCTACCCACTGCCCCAGCCTGTGCGGAGCTGAGCCAGCCTG 600
Db 541 TGTATGGCGACGCTTGGCTACCCACTGCCCCAGCCTGTGCGGAGCTGAGCCAGCCTG 600

Qy 601 GGGCCCTGGCGCTGCCACAGTACTTCTCCAGAGATGATGACTTCTGGCTGCTGAA 660
Db 601 GGGCCCTGGCGCTGCCACAGTACTTCTCCAGAGATGATGACTTCTGGCTGCTGAA 660

Qy 661 GGAGCTGCAGACTGGCTGTATGGCTTACGCAAGGACTTCAACCGGCTTAAGAGAGAT 720
Db 661 GGAGCTGCAGACTGGCTGTATGGCTTACGCAAGGACTTCAACCGGCTTAAGAGAGAT 720

Qy 721 GCAGCTCCAGCAGCTTCACTGACCCCTGCACTTGGAGCAGATGTTTCTGACCTCTGAC 780
Db 721 GCAGCTCCAGCAGCTTCACTGACCCCTGCACTTGGAGCAGATGTTTCTGACCTCTGAC 780

Qy 781 CCTTAACCCCCACACCTCCAGGCGCCAGTCACTGTGCTT 819
Db 781 CCTTAACCCCCACACCTCCAGGCGCCAGTCACTGTGCTT 819
```

```

QY 601 GGCCCTGCGCCCTGCCCCCAGTGACTTCTCCAGAGATGGATGATCTTGGCTGCTGAA 660
Db 601 GGCCCTGCGCCCTGCCCCCAGTGACTTCTCCAGAGATGGATGATCTTGGCTGCTGAA 660
QY 661 GGAGCTGCAGACCTGGCTATGCGGTTCCAGCAAGAGATTCACCCGCTTAAGAAGAT 720
Db 661 GGAGCTGCAGACCTGGCTATGCGGTTCCAGCAAGAGATTCACCCGCTTAAGAAGAT 720
QY 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGACATGGTTTCTGACCTCTGAC 780
Db 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGACATGGTTTCTGACCTCTGAC 780
QY 781 CTTTAAACCCACACCTCCAGGCCAGTCACTGCTGCTT 819
Db 781 CTTTAAACCCACACCTCCAGGCCAGTCACTGCTGCTT 819

RESULT 3
AF176913 819 bp mRNA linear ROD 04-OCT-1999
LOCUS Mus musculus neurotrophin-1/B-cell stimulating factor-3 mRNA,
DEFINITION complete cds.
ACCESSION AF176913
VERSION AF176913.1 GI:6007644
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
REFERENCE
AUTHORS Senaldi,G., Varnum,B.C., Sarmiento,U., Starnes,C., Lile,J.,
Scully,S., Guo,J., Elliott,G., McNinch,J., Shaklee,C.L.,
Freeman,D., Manu,F., Simonet,W.S., Boone,T. and Chang,M.S.
TITLE Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
IL-6 family
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)
MEDLINE 99432254
PUBMED 10500198
REFERENCE 2 (bases 1 to 819)
AUTHORS Senaldi,G., Varnum,B., Sarmiento,U., Lile,J., Starnes,C.,
Scully,S., Guo,J., Elliott,G., McNinch,J., Freeman,D., Shaklee,C.,
Manu,F., Simonet,S., Boone,T. and Chang,M.-S.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1999) Angen, Inc., One Angen Center Drive,
Thousand Oaks, CA 91320, USA
FEATURES
source Location/Qualifiers
1..819
/organism="Mus musculus"
/db_xref="taxon:10090"
95..772
/codon_start=1
/product="neurotrophin-1/B-cell stimulating factor-3"
/protein_id="AAP00993.1"
/db_xref="GI:6007643"
/translation="MDLRAGSGMGLACLTVMHLPAVPAALNRGDPGPGPSIQKTY
DLTRYELHQLRSAGTYLYNGPFPNPPRLGAEITPRATVNVWRSLNRLR
LTQNEYSHLLCYLRGLNQAATFAELRSLAHFCTSLQGLGSLIAGVATLGLPLPQ
PLPQTEPAWAPGPAHSDFLQKMDDFLLKELQTLWRSAXDFNRLKKMQPPAASVTL
HLEAHGF"
BASE COUNT 156 a 288 c 218 g 157 t
ORIGIN
Query Match 100.0%; Score 819; DB 10; Length 819;
Best Local Similarity 100.0%; Pred. No. 3.4e-176;
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGCAGCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGCAGCTCTGGGAGA 60
QY 61 GGAGCGCGCCGCGCGCGCCCGCCAGCCCTCCGAGCGGAGGAGCTGCTG 120

```

```

Db 61 GGAGCGCGCCGCGCGCGCCCGCCAGCCCTCCGAGCGGAGGAGCTGCTG 120
QY 121 GGGGATGTTAGCTTGCCTATGCAAGTGTGTGGGACCTCCCTCGAGTGCAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTGCCTATGCAAGTGTGTGGGACCTCCCTCGAGTGCAGCTCTTAA 180
QY 181 TCGCAGCAGAGATCCAGGCCCTTGGCCCTTCCATCCAGAAAAAATCTATGACCTCAACCCGCTA 240
Db 181 TCGCAGCAGAGATCCAGGCCCTTGGCCCTTCCATCCAGAAAAAATCTATGACCTCAACCCGCTA 240
QY 241 CTTGAGAGATCAACTCCGAGCTTAGCTGGGACCTTCACTGAACTACCTGGGGGCCCTTTT 300
Db 241 CTTGAGAGATCAACTCCGAGCTTAGCTGGGACCTTCACTGAACTACCTGGGGGCCCTTTT 300
QY 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGACAGAACTCTGCCAGGGCCACCGT 360
Db 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGACAGAACTCTGCCAGGGCCACCGT 360
QY 361 CAATTGGAAGTGTGGGAGGCTCAATGACAGGCTGGGGCTGACCCAGAACTATGAGGC 420
Db 361 CAATTGGAAGTGTGGGAGGCTCAATGACAGGCTGGGGCTGACCCAGAACTATGAGGC 420
QY 421 GTACAGTCACTCTCTGTGTACTTGGCTGGGCTCAACCGTCAAGGCTGCCAGCTGAACT 480
Db 421 GTACAGTCACTCTCTGTGTACTTGGCTGGGCTCAACCGTCAAGGCTGCCAGCTGAACT 480
QY 481 CCGAGCTAGCTGGGCCCACTTCTGTACAGGCTCCAGGGCTGGGGCTGAGCCAGCTTGCAGG 540
Db 481 CCGAGCTAGCTGGGCCCACTTCTGTACAGGCTCCAGGGCTGGGGCTGAGCCAGCTTGCAGG 540
QY 541 TGTATGCGCAGCGTTGGCTTACCCACTGCCCCAGGCTCTGCCAGGAGCTTGAGCCAGCTG 600
Db 541 TGTATGCGCAGCGTTGGCTTACCCACTGCCCCAGGCTCTGCCAGGAGCTTGAGCCAGCTG 600
QY 601 GGCCCTGCGCCCTGCGGACAGTACTTCTCAGAAAGATGATGATCTTGGCTGCTGAA 660
Db 601 GGCCCTGCGCCCTGCGGACAGTACTTCTCAGAAAGATGATGATCTTGGCTGCTGAA 660
QY 661 GGAGCTGCAGACCTGGCTATGCGGTTTCAGCAAGAGCTTCAACCGCTTAAGAAGAT 720
Db 661 GGAGCTGCAGACCTGGCTATGCGGTTTCAGCAAGAGCTTCAACCGCTTAAGAAGAT 720
QY 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGACATGGTTTCTGACCTCTGAC 780
Db 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGACATGGTTTCTGACCTCTGAC 780
QY 781 CTTTAAACCCACACCTCCAGGCCAGTCACTGCTGCTT 819
Db 781 CTTTAAACCCACACCTCCAGGCCAGTCACTGCTGCTT 819

RESULT 4
AX205024 881 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 1 from Patent WO0155172.
DEFINITION AX205024
ACCESSION AX205024
VERSION AX205024.1 GI:15394259
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 881)
TITLE Elson,G., Gauchat,J.F., Plun-Favreau,H., Chevalier,S. and Gascan,H.
Isolated complex comprising a nnt-1 protein and in addition at
least a clif-1 protein and/or a scntfr_g(a) protein
JOURNAL Patent: WO 0155172-A 1 02-AUG-2001;
PIERRE FABRE MEDICAMENT (FR) ; INSTITUT NATIONAL DE LA SANTE ET DE
LA RECHERCHE MEDICALE (INSERM) (FR)
FEATURES Location/Qualifiers
source 1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```
CDS
174..851
/note="NNT-1"
/codon_start=1
/protein_id="CAC60175.1"
/db_xref="GI:15394260"
/translation="MDLRAGDSWGLACLCTVLWHLPAVPALNRTGDPGPGPSIQKTY
DLTRYLEHQLRSLAGTYLNYLGPPEPNPRIGAEITLPRATVDLEVRSLNDKLR
LTQNYEAYSHLLCYLRGLNRQAATAELRSLAHFCTSLQGLLGSITAGVMAALGYPLPQ
PLPGTEPTWTPGPAHSDFLQMDDFLLKELQTLWRSKDFNRLKKKQPPAAAVTL
HLGAHGF"
BASE COUNT 158 a 318 c 246 g 159 t
ORIGIN
Query Match 82.3%; Score 674; DB 6; Length 881;
Best Local Similarity 92.2%; Pred. No. 3.5e-143; Indels 1; Gaps 1;
Matches 721; Conservative 0; Mismatches 60;
QY 2 ATTATTAAAGCTTTCGCCGAGCGCGGCTCGCCCTCCCACTCCGACGACCTCTGGGAGAG 61
DB 82 ATTATTAAAGCTTTCGCCGAGCGCGGCTCGCCCTCCCACTCCGACGACCTCTGGGAGAG 141
QY 62 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121
DB 142 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 200
QY 122 GGGATGTTAGCTTGCCTATGACCGGTGCTGTGGACCTCCCTCGAGTGCAGCTCTTAAT 181
DB 201 GGGATGTTAGCTTGCCTATGACCGGTGCTGTGGACCTCCCTCGAGTGCAGCTCTTAAT 260
QY 182 CGCACAGGAGATCCAGGCGCTCGCCCTCCATCCAGAAACCTATGACCTCACCCGCTAC 241
DB 261 CGCACAGGAGATCCAGGCGCTCGCCCTCCATCCAGAAACCTATGACCTCACCCGCTAC 320
QY 242 CTGGAGCATCACTCCGAGCTTAGCTGGACCTACTGAACCTACTTGGGGCGCCCTTTC 301
DB 321 CTGGAGCATCACTCCGAGCTTAGCTGGACCTACTGAACCTACTTGGGGCGCCCTTTC 380
QY 302 AACGAGCTGACTTCAATCTCTCGACTGGGGGAGAAACTCTGCCAGGCGCCACGGTC 361
DB 381 AACGAGCTGACTTCAATCTCTCGACTGGGGGAGAAACTCTGCCAGGCGCCACGGTC 440
QY 362 AACTTGAAGTGTGGGAGAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGCGG 421
DB 441 GACTTGAAGTGTGGGAGAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGCGG 500
QY 422 TACAGTCACTCTCTGTGTTACTTGGTGGCTCAACCGTGCAGCTGCAGCTGAATC 481
DB 501 TACAGTCACTCTCTGTGTTACTTGGTGGCTCAACCGTGCAGCTGCAGCTGAATC 560
QY 482 CGCAGTAGCTGGCGGCTTCTGTGTTACTTGGTGGCTCAACCGTGCAGCTGCAGCTGAAT 541
DB 561 CGCAGTAGCTGGCGGCTTCTGTGTTACTTGGTGGCTCAACCGTGCAGCTGCAGCTGAAT 620
QY 542 GTCATGGCAGCTTGGCTACCCACTGCCAGCTCTGCCAGGAGCTGAGCCAGCTGG 601
DB 621 GTCATGGCAGCTTGGCTACCCACTGCCAGCTCTGCCAGGAGCTGAGCCAGCTGG 680
QY 602 GCCCTTGGCGCTGCGCAGTACTTCTCTCGAAGATGATGATCTTGGCTGCTGAAG 661
DB 681 ACTCTTGGCGCTGCGCAGTACTTCTCTCGAAGATGATGATCTTGGCTGCTGAAG 740
QY 662 GAGCTGAGACCTTGGCTTATGGCTTACGCAAGACTTCAACCGGCTTAAAGAAAGATG 721
DB 741 GAGCTGAGACCTTGGCTTATGGCTTACGCAAGACTTCAACCGGCTTAAAGAAAGATG 800
QY 722 CAGCCTCCAGCAGCTTCACTCAGCTGAGTGGAGGACATGGTTTCTGACCTTGACC 781
DB 801 CAGCCTCCAGCAGCTTCACTCAGCTGAGTGGAGGACATGGTTTCTGACCTTGACC 860
QY 782 CT 783
DB 861 TT 862
```

```
QY 542 GTATGCGGACGCTTGCTGCTACCCAGCTCCCGAGCCTCTGCCAGGAGTGGAGCCAGCCTGG 601
Db 621 GTATGCGGACGCTTGCTGCTACCCAGCTCCCGAGCCTCTGCCAGGAGTGGAGCCAGCCTGG 680
QY 602 GCCCTTGGCCCTGCCACAGTACTTCTCCAGAGATGATGATGATGATGATGATGATGATGATG 661
Db 681 ACTCTTGGCCCTGCCACAGTACTTCTCCAGAGATGATGATGATGATGATGATGATGATGATG 740
QY 662 GAGCTGAGACCTGGCTATGGCTTTCAGCAAGGACTTCAACCGGCTTAAAGAGAGATG 721
Db 741 GAGCTGAGACCTGGCTATGGCTTTCAGCAAGGACTTCAACCGGCTTAAAGAGAGATG 800
QY 722 CAGCTCCAGCAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781
Db 801 CAGCTCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 860
QY 782 CT 783
Db 861 TT 862

RESULT 6
LOCUS AR002595 797 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5741772.
ACCESSION AR002595
VERSION AR002595.1 GI:3964149
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 797)
AUTHORS Chang, M.-S.
TITLE Neurotrophic factor NNT-1
JOURNAL Patent: US 5741772-A 1 21-APR-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 139 a 297 c 218 g 143 t
ORIGIN

Query Match 81.7%; Score 669.4; DB 6; Length 797;
Best Local Similarity 92.0%; Pred. No. 3.9e-142;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 5 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCCGAGCCTCTGGAGAGGAG 64
Db 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCCGAGCCTCTGGAGAGGAG 60
QY 65 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
Db 61 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
QY 125 ATGTTAGCTTGCTATGACAGCTGCTGCGCAGCTCCCTGCGAGTGCAGCTCTTAATCGC 184
Db 120 ATGTTAGCTTGCTATGACAGCTGCTGCGCAGCTCCCTGCGAGTGCAGCTCTTAATCGC 179
QY 185 ACAGAGATTCAGGCGCTTGGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTG 244
Db 180 ACAGGAGACCCAGGCGCTTGGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTG 239
QY 245 GAGCATCAACTCCGAGCTTAGCTGGAGCCTACCTGAACTACCTGGGCGCCCTTTCAC 304
Db 240 GAGCACCAGCTCCGAGCTTGGCTGGAGCCTATCTGAACTACCTGGGCGCCCTTTCAC 299
QY 305 GAGCTGACTTCAATCTCTCTGAGCTGGGGGAGAGAACTTGGCCAGAGAACTTGGCCAGAGCTCAAC 364
Db 300 GAGCAGACTTCAACCTCTCCCGGCTTGGGGGAGAGAACTTGGCCAGAGAACTTGGCCAGAGCTCAAC 359
QY 365 TTGGAAGTGTGGGAGGCTCAATGACAGGCTTGGGCTGAGCCAGAACTATGAGGCGCTAC 424
Db 360 TTGAGAGTGTGGGAGGCTCAATGACAACTGCGGCTGAGCCAGAACTATGAGGCGCTAC 419
```

```
QY 425 AGTCACCTCTGTGTACTTTCGCTGGCTCAACCGTCAGGCTGCCACAGCTGAATCCGA 484
Db 420 AGCACCTTCTGTGTACTTTCGCTGGCTCAACCGTCAGGCTGCCACAGCTGAATCCGA 479
QY 485 CGTAGCTGCGCCACCTTCTGTACAGCTCCAGGCTCGTGGGAGCTGCTGGGAGCTGCTGGGCGTC 544
Db 480 CGCAGCTGCGCCACCTTCTGTACAGCTCCAGGCTCGTGGGAGCTGCTGGGAGCTGCTGGGCGTC 539
QY 545 ATGGCAGCGCTTGGCTTACCCACTGCCCCAGCCTCTGCGAGGACTGAGCCAGCCTGGGCC 604
Db 540 ATGGCAGCTTCTGGGCTACCCACTGCCCCAGCCTCTGCGAGGACTGAGCCAGCCTGGGACT 599
QY 605 CTGGGCGCTGCGCCACAGTACTTCTCCAGAGATGATGATGATGATGATGATGATGATGATG 664
Db 600 CTGGGCGCTGCGCCACAGTACTTCTCCAGAGATGATGATGATGATGATGATGATGATGATG 659
QY 665 CTGCAGACCTGGCTATGGGCTTCAGCAAGGACTTCAACCGGCTTAAAGAGAGATGAG 724
Db 660 CTGCAGACCTGGCTATGGGCTTCAGCAAGGACTTCAACCGGCTTAAAGAGAGATGAG 719
QY 725 CTTCAGCAGCTTCACTGACCTGCACTTGGAGGCACTGGTTCCTGACCTCTGACCT 783
Db 720 CTTCAGCAGCTTCACTGACCTGCACTTGGAGGCACTGGTTCCTGACCTCTGACCT 778

RESULT 7
LOCUS AX392086 797 bp DNA linear PAT 23-MAR-2002
DEFINITION Sequence 1 from Patent WO0215977.
ACCESSION AX392086
VERSION AX392086.1 GI:19700574
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Senaldi, G.
TITLE Methods and compositions for treating ige-related disease using nn
JOURNAL t-1 inhibitors
FEATURES
source Location/Qualifiers
BASE COUNT 139 a 297 c 218 g 143 t
ORIGIN

Query Match 81.7%; Score 669.4; DB 6; Length 797;
Best Local Similarity 92.0%; Pred. No. 3.9e-142;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 5 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCCGAGCCTCTGGAGAGGAG 64
Db 1 ATTAAGCTTCGCGAGCGCGGCTCGCCCTCCCACTCCGAGCCTCTGGAGAGGAG 60
QY 65 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
Db 61 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
QY 125 ATGTTAGCTTGCTATGACAGCTGCTGCGCAGCTCCCTGCGAGTGCAGCTCTTAATCGC 184
Db 120 ATGTTAGCTTGCTATGACAGCTGCTGCGCAGCTCCCTGCGAGTGCAGCTCTTAATCGC 179
QY 185 ACAGAGATTCAGGCGCTTGGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTG 244
Db 180 ACAGGAGACCCAGGCGCTTGGCCCTCCATCCAGAAAACCTATGACCTCACCGGCTACCTG 239
QY 245 GAGCATCAACTCCGAGCTTAGCTGGAGCCTACCTGAACTACCTGGGCGCCCTTTCAC 304
Db 240 GAGCACCAGCTCCGAGCTTGGCTGGAGCCTATCTGAACTACCTGGGCGCCCTTTCAC 299
QY 305 GAGCTGACTTCAATCTCTCTGAGCTGGGGGAGAGAACTTGGCCAGAGAACTTGGCCAGAGCTCAAC 364
Db 300 GAGCAGACTTCAACCTCTCCCGGCTTGGGGGAGAGAACTTGGCCAGAGAACTTGGCCAGAGCTCAAC 359
QY 365 TTGGAAGTGTGGGAGGCTCAATGACAGGCTTGGGCTGAGCCAGAACTATGAGGCGCTAC 424
Db 360 TTGAGAGTGTGGGAGGCTCAATGACAACTGCGGCTGAGCCAGAACTATGAGGCGCTAC 419
```

Qy	125	ATGTTAGCTTGCTATGCA	CGGTGCTGTGGCACCTCCCTCGAGTGCAGCTCTTAATCGC	181
Db	120	ATGTTAGCGTGCCTGTGCA	CGGTGCTCTGGCACCTCCCTCGAGTGCAGCTCTTAATCGC	179
Qy	185	ACAGGAGATCACGSCCTG	SCCTCCATCCAGAAACCTATGACCTCACCCGCTACCTG	244
Db	180	ACAGGGAGCCAGGGCTG	CCCTCATCCAGAAACCTATGACCTCACCCGCTACCTG	239
Qy	245	GAGCATCAACTCCGCA	GCTTAGCTGGGAACCTACCTGAACTACCTGGGGCCCTTTCAAC	304
Db	240	GAGCACCAACTCCGCA	GCTTTGGTGGGACCTATCTGAACCTACCTGGGGCCCTTTCAAC	299
Qy	305	GAGCCTGACCTTCAAT	CTCTCTCGACTGGGGGCAAGAACTCTGCCAGGGGCAAGGTCACAC	364
Db	300	GAGCAGACTTTCAAC	CTCTCCCGCTGGGGGCAAGAACTCTGGCCAGGGGCAAGGTCACAC	359
Qy	365	TTTGAAGCTGTGGGAA	AGCCTCAATGACAGGCTCGCGGTGACCCAGAGAACTATGAGGCGTAC	424
Db	360	TTTGAAGCTGTGGGAA	AGCCTCAATGACAAACTGGGCTGACCCAGAGAACTATGAGGCGTAC	419
Qy	425	AGTCACCTCTCTGTG	TGTTACTTTGCGTGGGCTCAACCGCTCAGGCTGCCACAGCTGAACTCCGA	484
Db	420	AGCCACCTTCTGTG	TGTTACTTTGCGTGGGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGGCG	479
Qy	485	CGTAGCTGGCGCCACT	TTCTGTACACAGCTCTCAGGGCTGCTGGGCAAGCATTTGACAGTGTC	544
Db	480	CGCAGCTGGCGCCACT	TTCTGTGACACAGCTCTCAGGGCTGCTGGGCAAGCATTTGCGGGCGTC	539
Qy	545	ATGGCGACGCTTGCTG	CTACCACTGCCCCAGCCTCTGCCAGGGACTGAGCCAGCTCTGGGCC	604
Db	540	ATGGCAGCTCTGGCT	ATCCCACTGCCCACTGCCCGCTGGGACTGAAACCACTTGGACT	599
Qy	605	CTGGCCCTGCCACAG	TGACTTCTCCAGAAAGTGAATGACTTCTGGCTGCTGTAAGAGGAG	664
Db	600	CTGGCCCTGCCACAG	TGACTTCTCCAGAAAGTGAACACTTCTGGCTGCTGTAAGAGGAG	659
Qy	665	CTGCAGACCTGGCT	TATGGCGTTACGCCAAGGACTTTCAACCGGTTTAAGAAAGAGATGCAG	724
Db	660	CTGCAGACCTGGCT	TGCGGCTCGGCCAAGGACTTTCAACCGGCTCAAGAAAGAGATGCAG	719
Qy	725	CCTCCAGAGCTTCAG	TCACTCCGTGCACTTTGAGGAGCAATGGTTTCTGACCTCTGACCCCT	783
Db	720	CCTCCAGAGCTTCAG	TCACTCCGTGCACTCGGGGCTCATGGCTTCTGACTCTGACCTT	778

RESULT 8	AF176911	797 bp	mRNA	linear	PRI 04-OCT-1999
LOCUS	AF176911				
DEFINITION	Homo sapiens neurotrophin-1/B-cell stimulating factor-3 mRNA, complete cds.				
ACCESSION	AF176911				
VERSION	AF176911.1	GI:6007640			
KEYWORDS	.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Snaldi,G., Varnum,B.C., Sarmiento,U., Starnes,C., Lile,J., Scully,S., Guo,J., Elliott,G., McNinch,J., Shaklee,C.L., Freeman,D., Manu,F., Simonet,W.S., Boone,T. and Chang,M.S.				
TITLE	Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the lu-6 family				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11458-11463 (1999)				
MEDLINE	99432254				
PUBMED	10500198				
REFERENCE	2 (bases 1 to 797)				
AUTHORS	Snaldi,G., Varnum,B., Sarmiento,U., Lile,J., Starnes,C., Scully,S., Guo,J., Elliott,G., McNinch,J., Freeman,D., Shaklee,C., Manu,F., Simonet,S., Boone,T. and Chang,M.-S.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-AUG-1999) Amgen. Inc., One Amgen Center Drive,				

```

RESULT 9
BC012939
LOCUS
DEFINITION
Homo sapiens, similar to cardiotrophin-like cytokine;
neurotrophin-1/B-cell stimulating factor-3, clone MGC:21195
IMAGE:4453813, mRNA, complete cds.
ACCESSION
BC012939
VERSION
GI:15277894
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1736)
Direct Submission
Submitted (20-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalobos@bcm.tmc.edu.
Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov
Series: IRAP Plate: 28 Row: b Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6007640.
FEATURES
source
1..1736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:21195 IMAGE:4453813"
/tissue type="Kidney, hypernephroma"
/clone.lib="NIH_MGC_89"
/lab host="DH10B"
/note="vector: pCMV-SPORT6"
73..750
/codon_start=1
/product="Similar to cardiotrophin-like cytokine;
neurotrophin-1/B-cell stimulating factor-3"
/protein_id="AAH12939.1"
/db_xref="GI:15277895"
/translation="MDLRGDSWGMGLACTVLVHLPAVPAALNRITGDPGSPSIQKTY
LTRYLEHQLRSLAGTYLYNLPFPNPFNPRFPAETLPVATVLEWVRSNDKLR
LTQNYRAYSLCYLRGLNRQAELRSLAHSFCTSLQGLLGIAGVMAALGVPLPQ
PLFETPTWTPGPAHSDFLQKMDDFLLKELQTLWRSKDFNLKXKQMPFAAVTL
HLGAGGF"
BASE COUNT 371 a 546 c 454 g 365 t
ORIGIN
Query Match 79.9%; Score 654; DB 9; Length 1736;
Best Local Similarity 92.0%; Pred. No. 1.1e-138;
Matches 701; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
QY 22 GCGCGGCTCGCCCTCCCACTCCGCCAGCTCTGGGAGAGAGCGCGCGCGCGCC 81
Db 1 GCGCGGCTCGCCCTCCCACTCCGCCAGCTCTGGGAGAGAGCGCGCGCGCGCC 60

```

```

QY 82 GGCCCCCAGCCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGGCCTATG 141
Db 61 GG-CCCAAGCCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGGCCTATG 119
QY 142 CACGGTGTGTGGACCTCCCTGCACTGCCAGCTCTTAATCGCACAGGAGATCCAGGCC 201
Db 120 CACGGTGTGTGGACCTCCCTGCACTGCCAGCTCTTAATCGCACAGGAGATCCAGGCC 179
QY 202 TGGCCCCCTCCATCCAGAAAACCTATGACCTCACCCGCTACCTGGAGGATCAACTCCGAG 261
Db 180 TGGCCCCCTCCATCCAGAAAACCTATGACCTCACCCGCTACCTGGAGGATCAACTCCGAG 239
QY 262 CTTAGCTGGGACCTACTGAACTTACCTGGGGCCCCCTTTCAACGAGCCTGACTTTCAATCC 321
Db 240 CTTAGCTGGGACCTACTGAACTTACCTGGGGCCCCCTTTCAACGAGCCTGACTTTCAATCC 299
QY 322 TCTTGACCTGGGGGCGAGAAACTCTGCCCAGGGCCACGGTCAACTTTGGAAAGTGTGGGAAG 381
Db 300 TCCCCGCTGGGGGCGAGAACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAG 359
QY 382 CCTCAATGACAGGCTGGGCTGACCCAGAACTATGAGGGCTACAGTCACCTCTCTGTGTTA 441
Db 360 CCTCAATGACAACTGGGCTGACCCAGAACTATGAGGGCTACAGTCACCTCTCTGTGTTA 419
QY 442 CTTGCTGGCCCTCAACCGTCAAGCTGCCCAGCTGAACTCCGACGTAGCTTGGCCCCACTT 501
Db 420 CTTGCTGGCCCTCAACCGTCAAGCTGCCCAGCTGAACTCCGACGTAGCTTGGCCCCACTT 479
QY 502 CTGTACCAAGCTTCCAGGGCTCTGTCGGCAGCATTTGAGGTGTCTATGGCGACCTTGGCTA 561
Db 480 CTGCACCAAGCTTCCAGGGCTCTGTCGGCAGCATTTGAGGTGTCTATGGCGACCTTGGCTA 539
QY 562 CCCACTGCCCCAGCTCTGCCAGGAGCTGAGCCAGCTGAGCCCTGGGCCCTTGGCCCCACAG 621
Db 540 CCCACTGCCCCAGCTCTGCCAGGAGCTGAGCCAGCTGAGCCCTTGGAGCTTGGCCCCACAG 599
QY 622 TGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTATG 681
Db 600 TGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTATG 659
QY 682 GCGTTCAGCCAGGACTTCAACCGCTTAAAGAGAGATGAGCCCTCCAGCAGCTTCACT 741
Db 660 GCGCTCGGCCAAGGACTTCAACCGCTTAAAGAGAGATGAGCCCTCCAGCAGCTTCACT 719
QY 742 CACCTGCACTTGGAGGCACATGTTTCTGACCTCTGACCTT 783
Db 720 CACCTGCACTTGGAGGCATGTTTCTGACCTTCTGACCTTCTGACCTT 761

```

```

RESULT 10
AF172854
LOCUS
DEFINITION
Homo sapiens cardiotrophin-like cytokine c1c mRNA, linear
ACCESSION
AF172854
VERSION
AF172854.1 GI:5852980
KEYWORDS
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1689)
Shi, Y., Wang, W., Yourey, P.A., Gohari, S., Zukauskas, D., Zhang, J.,
Ruben, S. and Alderson, R.F.
Computational EST database analysis identifies a novel member of
the neuropoietic cytokine family
Biochem. Biophys. Res. Commun. 262 (1), 132-138 (1999)
99382254
PUBMED
10448081
REFERENCE
2 (bases 1 to 1689)
AUTHORS
Shi, Y.
Direct Submission
Submitted (28-JUL-1999) Molecular Biology, Human Genome Science,
Inc., 9410 Key West Avenue, Rockville, MD 20850, USA

```


769 CTGACCTCTGACCT 783


```

QY 513 TCAGGCGCTGCTGGGAGGATTCAGGTCATGGCGAGCGCTTGGCTACCCAGTGCCTCC 572
Db 421 TCCAGGCGCTGCTGGGAGGATTCAGGTCATGGCGAGCGCTTGGCTACCCAGTGCCTCC 480
QY 573 AGCTCTGCCAGGAGTGCAGCGCTGGGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 632
Db 481 AGCGCTGCTGGGAGTGCAGCGCTGGGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 540
QY 633 AGAAGATGGATGACTTCTGCTGCTGAAGGAGCTGCAGAGCTGCGCTGATGGCTTCAGCCA 692
Db 541 AGAAGATGGAGACTTCTGCTGCTGAAGGAGCTGCAGAGCTGCGCTGATGGCTTCAGCCA 600
QY 693 AGACTTCACCGGCTTAAGAAGATGAGAGCTGCAGAGCTTCAGTACCTGCACT 752
Db 601 AGGACTTCAACCGGCTCAAGAAGATGAGAGCTGCAGAGCTTCAGTACCTGCACT 660
QY 753 TGGAGGCACATGTTTCTGA 772
Db 661 TGGGGGCTATGGCTTCTGA 680

RESULT 12
AC109138
LOCUS
DEFINITION Mus musculus clone RP23-41B18, WORKING DRAFT SEQUENCE, 7 ordered
pieces.
ACCESSION AC109138
VERSION AC109138.4 GI:22381069
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 269355)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
TITLE Mus musculus, clone RP23-41B18
REFERENCE 2 (bases 1 to 269355)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepey,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fargo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 269355)
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barn,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepey,Y., Collymore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fargo,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,

```

```

TITLE
JOURNAL
COMMENT
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L14003
Center clone name: 41_B_18
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 267668 bases at least Q40
Consensus quality: 268459 bases at least Q30
Consensus quality: 268627 bases at least Q20
Insert size: 242000; agarose-fp
Insert size: 268755; sum-of-contigs
Quality coverage: 10.4 in Q20 bases; agarose-fp
Quality coverage: 9.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 100642: contig of 100642 bp in length
* 100643 100742: gap of 100 bp
* 100743 104988: contig of 4246 bp in length
* 104989 105088: gap of 100 bp
* 105089 112582: contig of 7494 bp in length
* 112583 112682: gap of 100 bp
* 112683 142041: contig of 29359 bp in length
* 142042 142141: gap of 100 bp
* 142142 179941: contig of 37800 bp in length
* 179942 180041: gap of 100 bp
* 180042 259317: contig of 79276 bp in length
* 259318 259417: gap of 100 bp
* 259418 269355: contig of 9938 bp in length.
FEATURES
source
Location/Qualifiers
1..269355
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-41B18"
/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature
1..100642
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
100743..104988
/notes="assembly_fragment"
105089..112582
/notes="assembly_fragment"
112683..142041
/notes="assembly_fragment"

```

McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:20451120.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L14003
Center clone name: 41_B_18

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 267668 bases at least Q40
Consensus quality: 268459 bases at least Q30
Consensus quality: 268627 bases at least Q20
Insert size: 242000; agarose-fp
Insert size: 268755; sum-of-contigs
Quality coverage: 10.4 in Q20 bases; agarose-fp
Quality coverage: 9.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 100642: contig of 100642 bp in length
* 100643 100742: gap of 100 bp
* 100743 104988: contig of 4246 bp in length
* 104989 105088: gap of 100 bp
* 105089 112582: contig of 7494 bp in length
* 112583 112682: gap of 100 bp
* 112683 142041: contig of 29359 bp in length
* 142042 142141: gap of 100 bp
* 142142 179941: contig of 37800 bp in length
* 179942 180041: gap of 100 bp
* 180042 259317: contig of 79276 bp in length
* 259318 259417: gap of 100 bp
* 259418 269355: contig of 9938 bp in length.

FEATURES
source
Location/Qualifiers
1..269355
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-41B18"
/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature
1..100642
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
100743..104988
/notes="assembly_fragment"
105089..112582
/notes="assembly_fragment"
112683..142041
/notes="assembly_fragment"

KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5087)

AUTHORS Chang, M.-s.
TITLE Neurotrophic factor NNT-1
JOURNAL Patent: US 5741772-A 3 21-APR-1998;
FEATURES Location/Qualifiers

source 1: 5087
/organism="unknown"

BASE COUNT 992 a 1746 c 1191 g 1158 t
ORIGIN

Query Match 52.5%; Score 429.6; DB 6; Length 5087;

Best Local Similarity 90.4%; Pred. No. 1.3e-87;

Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACCTACCTGGGCCCCCTTTCAACGAGCGCTGACTTCAATCCCTCTGAGCTGGGGG 335

Db 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCGAGACTTCAACCCCTCCCGCTGGGGG 3422

QY 336 CAGAACTCTGCCAGGCGCAAGCTCAACTTGGAGTGTGGCGAAGCTCAATGACAGGC 395

Db 3423 CAGAGACTCTGCCAGGCGCACTTTGACTTGGAGGTGTGGCGAAGCTCAATGACAAAC 3482

QY 396 TGGCGCTGACCCAGAACTATGAGGCGGTACAGTCACTCCCTGTGTACTTGGCGTGCCTCA 455

Db 3483 TGGCGCTGACCCAGAACTATGAGGCGGTACAGGCACTTCTGTGTACTTGGCGTGCCTCA 3542

QY 456 ACCGTCAAGGCTGCCACAGCTGAACCTCCGACGTAGCTTGGCCCACTTCTGTACCAAGCTCC 515

Db 3543 ACCGTCAAGGCTGCCACAGCTGAGCTGGCGCGAGCTTGGCCCACTTCTGTACCAAGCTCC 3602

QY 516 AGGCGCTGCTGGGCGAGCATTCAGAGTGTATGGCGAGCGCTTGGCTACCCACTGCCCGCAGC 575

Db 3603 AGGCGCTGCTGGGCGAGCATTCAGAGTGTATGGCGAGCTTGGCTACCCACTGCCCGCAGC 3662

QY 576 CTCTGCCAGGAGCTGAGCGAGCTGGGCGCTGGCCCTGGCCCTGCCACAGTGACTTCTCCAGA 635

Db 3663 CGCTGCTGGGAGCTGAACCCACTTGGACTCTGGCCCTGCCACAGTGACTTCTCCAGA 3722

QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCAGACTGGCTATGGCGTTTCAGCCAAAGG 695

Db 3723 AGATGGAGACTTCTGGCTGCTGAAGAGCTGCAGACTGGCTATGGCGTTTCAGCCAAAGG 3782

QY 696 ACTTCAACCGGCTTAAGAAGAGATGAGCGCTCCAGCAGCTTCACTCAACCTGCACTTGG 755

Db 3783 ACTTCAACCGGCTTAAGAAGAGATGAGCGCTCCAGCAGCTTCACTCAACCTGCACTTGG 3842

QY 756 AGGCACATGGTTTCTGACCTCTGACCCCT 783

Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

RESULT 15

AX392088

LOCUS

AX392088 5087 bp DNA linear PAT 23-MAR-2002

DEFINITION Sequence 3 from Patent WO0215977.

ACCESSION AX392088

VERSION AX392088.1 GI:19700576

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Senaldi, G.

TITLE Methods and compositions for treating ige-related disease using nn

JOURNAL t-1 inhibitors

Patent: WO 0215977-A 3 28-FEB-2002;

Amgen Inc. (US)

FEATURES Location/Qualifiers

source

1: 5087

/organism="Homo sapiens"

misc_feature 137-138

/db_xref="taxon:9606"

/note="INTERVENING UNSEQUENCED REGION OF >1KB"

BASE COUNT 992 a 1746 c 1191 g 1158 t

ORIGIN

Query Match

Best Local Similarity 52.5%; Score 429.6; DB 6; Length 5087;

Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACCTACCTGGGCCCCCTTTCAACGAGCGCTGACTTCAATCCCTCTGAGCTGGGGG 335

Db 3363 AGCTGAACCTACCTGGGCCCCCTTTCAACGAGCGAGACTTCAACCCCTCCCGCTGGGGG 3422

QY 336 CAGAACTCTGCCAGGCGCAAGCTCAACTTGGAGTGTGGCGAAGCTCAATGACAGGC 395

Db 3423 CAGAGACTCTGCCAGGCGCACTTTGACTTGGAGGTGTGGCGAAGCTCAATGACAAAC 3482

QY 396 TGGCGCTGACCCAGAACTATGAGGCGGTACAGTCACTCCCTGTGTACTTGGCGTGCCTCA 455

Db 3483 TGGCGCTGACCCAGAACTATGAGGCGGTACAGGCACTTCTGTGTACTTGGCGTGCCTCA 3542

QY 456 ACCGTCAAGGCTGCCACAGCTGAACCTCCGACGTAGCTTGGCCCACTTCTGTACCAAGCTCC 515

Db 3543 ACCGTCAAGGCTGCCACAGCTGAGCTGGCGCGAGCTTGGCCCACTTCTGTACCAAGCTCC 3602

QY 516 AGGCGCTGCTGGGCGAGCATTCAGAGTGTATGGCGAGCGCTTGGCTACCCACTGCCCGCAGC 575

Db 3603 AGGCGCTGCTGGGCGAGCATTCAGAGTGTATGGCGAGCTTGGCTACCCACTGCCCGCAGC 3662

QY 576 CTCTGCCAGGAGCTGAGCGAGCTGGGCGCTGGCCCTGGCCCTGCCACAGTGACTTCTCCAGA 635

Db 3663 CGCTGCTGGGAGCTGAACCCACTTGGACTCTGGCCCTGCCACAGTGACTTCTCCAGA 3722

QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCAGACTGGCTATGGCGTTTCAGCCAAAGG 695

Db 3723 AGATGGAGACTTCTGGCTGCTGAAGAGCTGCAGACTGGCTATGGCGTTTCAGCCAAAGG 3782

QY 696 ACTTCAACCGGCTTAAGAAGAGATGAGCGCTCCAGCAGCTTCACTCAACCTGCACTTGG 755

Db 3783 ACTTCAACCGGCTTAAGAAGAGATGAGCGCTCCAGCAGCTTCACTCAACCTGCACTTGG 3842

QY 756 AGGCACATGGTTTCTGACCTCTGACCCCT 783

Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

Search completed: February 1, 2003, 18:03:37

Job time : 2162.75 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 1663.54 Seconds
(without alignments)
7973.437 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 819

Sequence: 1 tattataaagcttcgcgg.....aggccagctcagctgtgctt 819

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	644.4	78.7	887	14	BQ948158
2	568.6	69.4	1157	14	BQ940483
3	544	66.4	1053	12	BG164929
4	543	66.3	853	12	BG437538
5	504.6	61.6	594	14	BM763333
6	491.2	60.0	580	14	BM848189

7	485.2	59.2	573	14	BM840863
8	476	58.1	476	10	BE632644
9	464.4	56.7	955	9	AL543945
10	447	54.6	522	14	BM841897
11	445	54.3	542	14	BM821005
12	441.4	53.9	529	14	BM847924
13	440.893	53.8	532	14	BM846370
14	407.6	49.8	569	14	BM846370
15	377.8	46.1	458	13	BM363136
16	351.2	42.9	420	14	BM764238
17	333.6	40.7	512	9	AI390475
18	329.8	40.3	482	14	BM846622
19	312.4	38.1	407	9	AI752561
20	263	32.1	291	10	BM864730
21	246.8	30.1	552	12	EG095271
22	246.893	30.1	488	9	AA015243
23	228.4	27.9	440	12	BG148676
24	228	27.8	913	13	BI912197
25	172.8	21.1	915	12	BF035982
26	155.8	19.0	560	9	AL800340
27	124.8	15.2	691	12	BF213570
28	111.8	13.7	633	13	BU099801
29	102.6	12.5	655	13	BU096761
30	66.4	8.1	180	9	AA204015
31	50.8	6.2	711	12	BG758365
32	50.2	6.1	925	17	CNS0091P
33	45.6	5.6	1309	17	AG077201
34	45	5.5	925	17	CNS0091P
35	43.4	5.3	853	14	BQ951739
36	43.4	5.3	936	17	CNS01SR2
37	43.4	5.3	1101	17	CNS01288
38	42.8	5.2	844	17	CNS0052P
39	42.8	5.2	932	17	CNS0072Q
40	42.4	5.2	505	9	AI984845
41	42.4	5.2	658	10	AW873554
42	42.4	5.2	685	13	BM005788
43	42.4	5.2	690	12	BF732971
44	42.4	5.2	787	12	BF971932
45	42.4	5.2	865	14	BQ672955

ALIGNMENTS

RESULT 1
BQ948158
LOCUS
DEFINITION
AGENCOURT_8813192 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:6428214
5', mRNA sequence.
ACCESSION
BQ948158
VERSION
BQ948158.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 887)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: LNCM2614 row: h column: 07
High quality sequence stop: 674.
Location/Qualifiers
1. .887

```
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:6428214"
/clone_lib="NIH_MGC_101"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 154 a 325 c 240 g 167 t 1 others
ORIGIN
Query Match 78.7%; Score 644.4; DB 14; Length 887;
Best Local Similarity 92.3%; Pred. No. 3.6e-142;
Matches 700; Conservative 0; Mismatches 56; Indels 2; Gaps 2;
QY 2 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGAG 61
Db 5 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCTCTGGGAGAG 64
QY 62 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121
Db 65 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 123
QY 122 GGGATGTTAGCTTGCTTATGTCAGCGTGTGTGGCAGCTCCCTCGAGTGCAGCTTTAAT 181
Db 124 GGGATGTTAGCTTGCTTATGTCAGCGTGTGTGGCAGCTCCCTCGAGTGCAGCTTTAAT 183
QY 182 CGCACAGGAGATCCAGCGCTGCGCCCTCCATCCAGAAAACTATGACCTCAACCGGTAC 241
Db 184 CGCACAGGAGATCCAGCGCTGCGCCCTCCATCCAGAAAACTATGACCTCAACCGGTAC 243
QY 242 CTGGAGCATCACTCCGAGCTTAGTGGGAGCTAGTCACTGAGTCACTGCGGCGCGCTTC 301
Db 244 CTGGAGCATCACTCCGAGCTTAGTGGGAGCTATCTGAACTACCTTGGGCGCGCTTC 303
QY 302 AACGAGCTGACTTCAATCTCTCTCGAC-TGGGGGCGAGAACTCTGCCAGGGCGACGGT 360
Db 304 AACGAGCTGACTTCAATCTCTCTCGAC-TGGGGGCGAGAACTCTGCCAGGGCGACGGT 363
QY 361 CAATTTGGAAGTGTGGGAGAGCTCAATGACAGCTCGGCTGACCCAGCAACTATGAGGC 420
Db 364 TGACTTGGAGTGTGGGAGAGCTCAATGACAACTCGGCTGACCCAGCAACTATGAGGC 423
QY 421 GTACAGTCACTCTCTGTGTTACTTGGTGGCTCAACCGTCAAGTGCACAGCTGAAC 480
Db 424 CTACAGCACTCTCTGTGTTACTTGGTGGCTCAACCGTCAAGTGCACAGCTGAAC 483
QY 481 CCGAGCTAGCTTGCGCCCACTTGTATACAGCTCCAGGGCTGTGGGAGCACTTGCAGG 540
Db 484 CGCGCGAGCTTGCGCCCACTTGTATACAGCTCCAGGGCTGTGGGAGCACTTGCAGG 543
QY 541 TGTGATGGAGAGCTTGCGCTACCACTGCGCCAGCTCTGCCAGGAGTGAACCCACTTG 600
Db 544 CGTATGGAGAGCTTGCGCTACCACTGCGCCAGCTCTGCCAGGAGTGAACCCACTTG 603
QY 601 GGCCCTTGCGCTCCCGCAGTGAATCTCTCCAGAGATGATGATCTTCTGGCTCTGAA 660
Db 604 GATCTTGCGCTCCCGCAGTGAATCTCTCCAGAGATGATGATCTTCTGGCTCTGAA 663
QY 661 GGAGCTCAGACTTGCGCTATGGGCTTACGCAAGGATTTCAACCGGCTTAAGAGAAAT 720
Db 664 GGAGCTCAGACTTGCGCTATGGGCTTACGCAAGGATTTCAACCGGCTTAAGAGAAAT 723
QY 721 GCAGCTCCAGAGCTTCAGTCAACCTGCACTTGAGG 758
Db 724 GCAGCTCCAGAGCTTCAGTCAACCTGCACTTGAGG 761
```

```
RESULT 2
BQ940483
LOCUS
DEFINITION BQ940483 1157 bp mRNA linear EST 21-AUG-2002
IMAGE:6197786 5', mRNA sequence.
ACCESSION BQ940483.1 GI:22355961
VERSION BQ940483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1157)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgraphs-k@mail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LLAM13608 row: g column: 03
High quality sequence stop: 572.
FEATURES
Location/Qualifiers
1..1157
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:6197786"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCCG-3' and
5'-GACTAGTCTAGATCGGAGCGCGCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 221 a 374 c 346 g 216 t
ORIGIN
Query Match 69.4%; Score 568.6; DB 14; Length 1157;
Best Local Similarity 89.8%; Pred. No. 3.4e-124;
Matches 622; Conservative 0; Mismatches 69; Indels 2; Gaps 1;
QY 109 AGGGAGCTCGTGGGGATGTTAGCTTGCCTATGACAGGTGCTTGGACCTCCCTGCAGT 168
Db 58 AGGGAGCTCGTGGGGATGTTAGCTTGCCTATGACAGGTGCTTGGACCTCCCTGCAGT 117
QY 169 GCAGCTCTTAATCCACAGGAGATCCAGCGCTGCGCCCTCCATCCAGAAAACTATGA 228
Db 118 GCAGCTCTTAATCCACAGGAGATCCAGGGCTGCGCCCTCCATCCAGAAAACTATGA 177
QY 229 CCTCACCGCTACCTGGAGCATCACTCCGAGCTTGTAGTGGGAGCTTACCTGAACCTACCT 288
Db 178 CCTCACCGCTACCTGGAGCATCACTCCGAGCTTGTAGTGGGAGCTTACCTGAACCTACCT 237
QY 289 GGGGGCCCCCTTTCACAGCAGCTGACTTCAATCTCTCTCGACTGGGGGAGAACTCTGCC 348
Db 238 GGGGGCCCCCTTTCACAGCAGCTGACTTCAACCTCTCCCGCTGGGGGAGAGACTCTGCC 297
QY 349 CAGGGCCACGGTCAACTTGGAGTGTGGGAGAGCTCAATGACAGGCTGGGGCTGAGCCCA 408
```

```

Db 298 CAGGGCCACTGTTGACTTGGAGGTGTGGCAAGGCTCAATGACAACTGCGGCTGACCCA 357
QY 409 GAACTATGAGCGGTACAGTCACTCTCTGTTACTTGGTGGCTCAACCTGAGGCTGC 468
Db 358 GAACTAGAGCGCTACAGCACCCTTCTGTACTTGGTGGCTCAACCTGAGGCTGC 417
QY 469 CACAGCTGAATCTCGAGTACGCTGCGCCACTTCTGTACAGGCTTCCAGGCGCTGCTGGG 528
Db 418 CACTGTGAGTGGCGCGAGCTTGGCCCACTTCTGTACACAGCTTCCAGGCGCTGCTGGG 477
QY 529 CAGCATTTGCAAGTGTATGGCGAGCTTGGCTACCCCTGAGCTTCCAGAGATGAGTACTT 648
Db 478 CAGCATTTGCGGGCTCATGCGAGCTTGGCTACCCCTGAGCTTCCAGAGATGAGTACTT 537
QY 589 TGAGCCAGCTTGGGCGCTTGGCTGCGCCCTGCGCCCAAGTACTTCTCCAGAGATGAGTACTT 648
Db 538 TGAACCCACTTGGACTCTGCGCCCTGCGCCCAAGTACTTCTCCAGAGATGAGTACTT 597
QY 649 CTGGCTGCTGAAGAGCTGAGAGCTGAGAGCTTGGCTATGGCGTTACGCCAAGAGTCAACCGGCT 708
Db 598 CTGGCTGCTGAAGAGCTGAGAGCTTGGCTATGGCGTTACGCCAAGAGTCAACCGGCT 657
QY 709 TAAGAGAGATGAGAGCTTCCAGAGCTTCAAGTACTTCTCCAGAGATGAGTACTT 766
Db 658 CAAGAGAGATGAGAGCTTCCAGAGCTTCAAGTACTTCTCCAGAGATGAGTACTT 717
QY 767 TTCTGACCTTGAACCTTCAACCCCAAGTCTCC 799
Db 718 TTCTGACCTTGAACCTTCAACCCCAAGTCTCC 750

RESULT 3
BG164929
LOCUS
DEFINITION 60234355F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4453813 5',
mRNA sequence.
ACCESSION BG164929
VERSION BG164929.1 GI:12671563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1053)
NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Cloning Sequencing by: Incyte Genomics, Inc.
Cloning Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10244 row: m column: 14
High quality sequence start: 3
High quality sequence stop: 675.
Location/Qualifiers
1..1053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4453813"
/clone_lib="NIH_MGC_89"
/tissue_type="hypermorphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: kidney; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 210 a 351 c 292 g 200 t

```

ORIGIN

```

Query Match 66.4%; Score 544; DB 12; Length 1053;
Best Local Similarity 88.2%; Pred. No. 2.2e-118;
Matches 660; Conservative 0; Mismatches 80; Indels 8; Gaps 6;

QY 24 CGCGGCTGCGCTCCCACTCGCCAGGCTTGGGAGAGGAGCGCGCGCGCGCGCGCG 83
Db 2 CGCGGCTGCGCTCCCACTCGCCAGGCTTGGGAGAGGAGCGCGCGCGCGCGCGCG 61
QY 84 CCCCCAGCCCCCATGGAGCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGCCTATGCA 143
Db 62 CCCCCAGCCCCCATGGAGCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCTTGCCTATGCA 121
QY 144 CGGTGCTGTGGCACTCCCTGCAAGTGCAGCTCTTAAATCGACAGAGATGTCAGGCGCTG 203
Db 122 CGGTGCTGTGGCACTCCCTGCAAGTGCAGCTCTTAAATCGACAGAGATGTCAGGCGCTG 181
QY 204 GCCCTTCCATCCAGAAACCTATGACCTCACTGCGGCGCTACCTGGAGCATCAACTCGCAGCT 263
Db 182 GCCCTTCCATCCAGAAACCTATGACCTCACTGCGGCGCTACCTGGAGCATCAACTCGCAGCT 241
QY 264 TAGCTGGGACCTACCTGAACTACCTGGGCGCGCTTTCATCAACGAGCTT-GACTTCAAT 319
Db 242 TGGCTGGGACCTATCTGAATAGCTTGGGCGCGCTTTCATCAACGAGCTTGAACCTTCAAC 301
QY 320 CTTCTTCGACTGGGCGCAGAACTCTG-CCAGGGCGCAGGTCAACTTGGAACTGTGGCG 378
Db 302 CTTCTTCGACTGGGCGCAGAACTCTG-CCAGGGCGCAGGTCAACTTGGAACTGTGGCG 361
QY 379 AAGCTCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGGGTACAGTCACTCTCTGTG 438
Db 362 AAGCTCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGGGTACAGTCACTCTCTGTG 421
QY 439 TTACTTGGGTGCGCTCAACCGTCAAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTTCTGTG 498
Db 422 TTACTTGGGTGCGCTCAACCGTCAAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTTCTGTG 481
QY 499 CTTCTGACAGGCTTCCAGGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTTCTGTG 558
Db 482 CTTCTGACAGGCTTCCAGGCGCTGCGGCTGCGGCTGCGGCTGCGGCTTCTGTG 541
QY 559 CTACCCACTG-CCCCAGGCTTCTGCCAGGAGCTG-AGCCAGGCTTGGGCGCGCTGCGGCTTCTGTG 616
Db 542 CTACCCACTGCCCCAGGCTTCTGCCAGGAGCTG-AGCCAGGCTTGGGCGCGCTTCTGTG 601
QY 617 CACAGTGAATCTTCTCCAGAGATGAGTACTTCTGGCTGCTGAAAGAGAGTGCAGACCTGG 676
Db 602 CACAGTGAATCTTCTCCAGAGATGAGTACTTCTGGCTGCTGAAAGAGAGTGCAGACCTGG 661
QY 677 CTATGGCGCTTCCAGCAGGAGCTTCAACCGGCTTAAAGAGAGATGACAGCTTCCAGAGCT 736
Db 662 CTGTGGCGCTGGCG-AGGACTTCCAGGCTTCCAGAGAGATGACAGCTTCCAGAGCT 720
QY 737 TCAGTCACTTCCAGGAGCTTGGAGGAGCATG 764
Db 721 GCAGTCACTTCCAGGAGCTTGGAGGAGCATG 748

RESULT 4
BG437538
LOCUS
DEFINITION 60248927F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4621530 5',
mRNA sequence.
ACCESSION BG437538
VERSION BG437538.1 GI:13344044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 853)
NIH-MGC http://mgi.nci.nih.gov/.

```


	Matches	545;	Conservative	0;	Mismatches	49;	Indels	1;	Gaps	1;
Qy	33	CCCTCCCACTCGCCAGAGCCTCTGGGAGAGGAGCGCGCCGCGCGCCGCGCCGCCCCAGCC	92							
Db	1	CTCTCCCACTCGCCAGCCTCTGGGAGAGGAGCGCACCGCGCGCGCCCGG-CCCCAGCC	59							
Qy	93	CCATGGACCTTCGAGCAGGGGACTCTGTGGGGGATGTTTAGCTTGCTATGCAACGGTGTGT	152							
Db	60	CCATGGACCTTCGAGCAGGGGACTCTGTGGGGGATGTTAGCTGCTGTGTGCACGGTGTCT	119							
Qy	153	GGCACTCTCCCTGCACTGCGACGCTCTTAATCCGACAGGAGATCGAGGCGCTGGCCCTCCA	212							
Db	120	GGCACTCTCCCTGCACTGCGACGCTCTTAATCGCAGGGGACCCAGGGCGCTGGCCCTCCA	179							
Qy	213	TCCAGAAAACCTATGACCTCA CCGCTACCTGGAGCATCAACTCCGCAAGCTTAGCTGGGA	272							
Db	180	TCCAGAAAACCTATGACCTCA CCGCTACCTGGAGCACCAACTCCGCAAGCTTAGCTGGGA	239							
Qy	273	CCTACCTGMACTACTCGGGGCCCCCTTTCACAGAGCCTGACTTCAATCTCTCTCGACTGG	332							
Db	240	CCTATCTGAACTACTCGGGGCCCCCTTTCACAGAGCAGACTTCAACCCCTCCCGCCTGG	299							
Qy	333	GGGCAGAAACTCTGCCCAGGGCCACCGCTCAACTTGGAGGTGTGCGAAGGCTCAATGACA	392							
Db	300	GGGCAGAGACTCTGCCCCAGGGCCACTGTTCGACTGGAGGTGTGGCGAAGGCTCAATGACA	359							
Qy	393	GGCTGGCGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGCGTGGCC	452							
Db	360	AACTGGCGGCTGACCCAGAACTAGAGGCGCTACAGCCACTTCTGTGTATTCTGGTGGCC	419							
Qy	453	TCAACCGTCAAGGTGCCACAGCTGAATCCGAGTAGCCTGGGCCCACTTCTGTATCAGACC	512							
Db	420	TCAACCGTCAAGGTGCCACCTGTCTGAGTGTGCGCGCAGCCTGGGCCCACTTCTGCACCA	479							
Qy	513	TCAGGGCCTGTCTGGGCAGCAITTCAGAGGTTCATGGCGACGCTTGCTACCCACTGCCCC	572							
Db	480	TCAGGGCCTGTCTGGGCAGCAITTCAGGGGCTCATGGCAGCTCTGGGCTACCCACTGCCCC	539							
Qy	573	AGCCCTGCGCAGGACTGACCGAGCCTGGGCCCCCTGGCCCTGCCCCACAGTGAATT	627							
Db	540	AGCCGCTGCCGTGGGACTGAAACCACTTGGAATCTCTGGCCCTGCCCCACAGTGAATT	594							

RESULT 6	BM848189	580 bp	mRNA	linear	EST 06-MAR-2002
LOCUS	BM848189				
DEFINITION	K-ES070127911 S13KWS5 Homo sapiens cDNA clone S13KWS5-48-H08 5', mRNA sequence.				
ACCESSION	BM848189				
VERSION	BM848189.1 GI:19204588				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 580)				
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.				
TITLE	21C Frontier Korea EST Project 2001				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 48 row: H column: 08 High quality sequence stop: 580.				

```

FEATURES
  source      Location/Qualifiers
              1. .580
              /organism="Homo sapiens"
              /quality sequence stop: 580.

```

```

/db_xref=taxon:9606"
/clone="S13KMS5-48-H08"
/clone_lib="S13KMS5"
/tissue_type="myeloma"
/cell_line="KMS-5"
/lab_host="Topi0P."
/notes=vector: pCNS; Site_1: EcoRI; Site_2: NotI; The poly
(A)+ RNA was dephosphorylated with bacterial alkaline
phosphatase (BAP) and then decapped with tobacco acid
pyrophosphatase (TAP). The decapped intact mRNA was
ligated with DNA-RNA linker including EcoR I site by
treatment of T4 RNA ligase and the first strand cDNA was
synthesized from oligo dt-selected mRNA by priming with
dt-tailed vector. The dt-tailed vector was adjusted to
have about 60nt. The cDNA vector was circularized with E.
coli DNA ligase after digestion of EcoRI which site is
also included in vector. An RNA strand converted to a DNA
strand by Okayama-Berg method. The obtained cDNA vectors
were used for transformation of competent cells E. coli
Topi0P. by electroporation method. The cDNA libraries
constructed by this method are full-length enriched cDNA
library."

```

[illegible]

DEFINITION K-EST0118015 S13KMS5 Homo sapiens cDNA clone S13KMS5-35-D11 5', mRNA sequence.

ACCESSION BMB40863

VERSION BMB40863.1 GI:19197272

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 35 row: D column: 11

High quality sequence stop: 573.

FEATURES

source

1..573

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S13KMS5-35-D11"

/clone_lib="S13KMS5"

/tissue_type="myeloma"

/cell_line="KMS-5"

/lab_host="Top10F"

/notes="Vector: pCNS; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 98 a 217 c 160 g 98 t

ORIGIN

Query Match 59.2%; Score 485.2; DB 14; Length 573;

Best Local Similarity 91.5%; Pred. No. 1.5e-104;

Matches 525; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 35 CTCCTCCCTCCGAGCGGACTCGTGGGGATGTTAGCTGTGCTATGACGGTGTGTGG 94

DB 1 CTCCTCCCTCCGAGCGGACTCGTGGGGATGTTAGCTGTGCTATGACGGTGTGTGG 59

QY 95 ATGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTGTGCTATGACGGTGTGTGG 154

DB 60 ATGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTGTGCTATGACGGTGTGTGG 119

QY 155 CACCTCCCTGAGTCCGAGCTTTAATCGACAGGATCCAGGCTCGGCTCCATC 214

DB 120 CACCTCCCTGAGTCCGAGCTTTAATCGACAGGATCCAGGCTCGGCTCCATC 179

QY 215 CAGAAACCTATGACTCACCGGCTACCTGGAGCATCACTCCGAGCTTAGTGGGACC 274

DB 180 CAGAAACCTATGACTCACCGGCTACCTGGAGCATCACTCCGAGCTTAGTGGGACC 239

QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAACAGAGCTGACTTCATCTCTCGACTGGGG 334

Db 240 TATCTGAACCTACCTGGGCCCTTTTCAACGAGCAGACTTCAACCTCCCGCTGGG 299

QY 335 GCAGAAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGCGAAGCTCAATGACAGG 394

Db 300 GCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGCGAAGCTCAATGACAAA 359

QY 395 CTGGGCTGACCCAGAACTATGAGCGGTACAGTCACTCTCTGTGTACTTGGTGGCTTC 454

Db 360 CTGGGCTGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGTGGCTTC 419

QY 455 AACCGTCAGGCTGCCAGCTGAACTCCGAGCTAGCTGGCCCTTCTGTGTACTTGGTGGCTTC 514

Db 420 AACCGTCAGGCTGCCAGCTGAACTCCGAGCTAGCTGGCCCTTCTGTGTACTTGGTGGCTTC 479

QY 515 CAGGCTCTGGGCGAGCATTTGCGGCGCTATGCGGCGCTTGGCTACCCACTGGCCCCAG 574

Db 480 CAGGCTCTGGGCGAGCATTTGCGGCGCTATGCGGCGCTTGGCTACCCACTGGCCCCAG 539

QY 575 CCTCTGCCAGGACTGAGCCAGCTGGGCCCTTGGGCCCTG 608

Db 540 CCGCTGCTGGGACTGAACCCACTTGGACTCCTG 573

RESULT 8

BE632644

LOCUS uv63f01.yl Soares mouse 476 bp mRNA linear EST 25-AUG-2000

DEFINITION 5' mRNA sequence.

ACCESSION BE632644

VERSION BE632644.1 GI:9915332

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 476)

AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1087677

Seq primer: -40RP from Gibco

High quality sequence stop: 464.

Location/Qualifiers

1..476

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:3411865"

/clone_lib="Soares mouse 3NBMS"

/sex="male"

/tissue_type="Spleen"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TTTTACCAATCTGAAGTGGAGCGGCGGCTGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 93 a 165 c 126 g 92 t

ORIGIN

Query Match 58.1%; Score 476; DB 10; Length 476;

```
Best Local Similarity 100.0%; Pred. No. 2.1e-102;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GSCCCTCCATCCAGAAACCTATGACCTCACCCGCTACCTGGAGCATCAACTCCGACG 262
D 1 GSCCCTCCATCCAGAAACCTATGACCTCACCCGCTACCTGGAGCATCAACTCCGACG 60
QY 263 TTAGTGGGACCTACTGAACTACTCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCT 322
D 61 TTAGTGGGACCTACTGAACTACTCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCT 120
QY 323 CTTGACTGGGGGAGAACTCTGCCAGGGCCAGCTCACTGGAAGTGGGGAGG 382
D 121 CTTGACTGGGGGAGAACTCTGCCAGGGCCAGCTCACTGGAAGTGGGGAGG 180
QY 383 CTCAATGACAGCTCGGCTGACCCAGAACTATGAGGGGTACAGTCACCTCTGTGTTAC 442
D 181 CTCAATGACAGCTCGGCTGACCCAGAACTATGAGGGGTACAGTCACCTCTGTGTTAC 240
QY 443 TTGGTGGCCCTCAACCTGACGCTGCCAGCTGAATCCGAGCTGAGCTGCCCACTTC 502
D 241 TTGGTGGCCCTCAACCTGACGCTGCCAGCTGAATCCGAGCTGAGCTGCCCACTTC 300
QY 503 TGTACAGCTCCAGGGCTGTCTGGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 562
D 301 TGTACAGCTCCAGGGCTGTCTGGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 360
QY 563 CCATGCCCCAGCTCTGCCAGGACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 622
D 361 CCATGCCCCAGCTCTGCCAGGACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 420
QY 623 GACTTCTCCAGAGATGATGACTTCTGGTGTGTAAGAGAGCTGACAGCTGGCT 678
D 421 GACTTCTCCAGAGATGATGACTTCTGGTGTGTAAGAGAGCTGACAGCTGGCT 476

RESULT 9
AL543945
LOCUS
DEFINITION AL543945 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1004YML5 5
prime, mRNA sequence.
VERSION AL543945
KEYWORDS AL543945.1 GI:12876424
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
COMMENT Full-length cDNA libraries and normalization
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. .955
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="CS0D1004YML5"
/tissue="placenta"
/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
```

```
BASE COUNT 168 a 323 c 266 g 198 t
ORIGIN
Query Match 56.7%; Score 464.4; DB 9; Length 955;
Best Local Similarity 90.5%; Pred. No. 1.5e-99;
Matches 507; Conservative 0; Mismatches 51; Indels 2; Gaps 1;

QY 224 TATGACCTCACCCGCTACCTGGAGCATCAACTCCGAGCTTACCTGGGACCTACCTGAAC 283
D 1 TATGACCTCACCCGCTACCTGGAGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAAC 60
QY 284 TACTTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCTCTCTCGACTGGGGGAGAACT 343
D 61 TACTTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGGGAGAGACT 120
QY 344 CTGCCAGGGCCAGGTCACCTTGAAGTGTGGGAGAGCTCAATGACAGGCTGGGGCTG 403
D 121 CTGCCAGGGCCAGGTCGTTGACTTGGAGGTGGGCGAA--CTCAATGACAAACTGGGGCTG 178
QY 404 ACCAGAACTATGAGGCGTACAGTCACCTCTCTGTACTTGTGCTGGCTCAACCGTCAG 463
D 179 ACCAGAACTATGAGGCGCTACAGCCACCTTCTGTGTACTTGGCTGAGCTCAACCGTCAG 238
QY 464 GCTGCCAGCTGAATCCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 523
D 239 GCTGCCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 298
QY 524 CTGGGAGCATTTGAGGCTGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 583
D 299 CTGGGAGCATTTGAGGCTGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 358
QY 584 GGGACTGAGCCAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 643
D 359 GGGACTGAGCCAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 418
QY 644 GACTTCTGGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAG 703
D 419 GACTTCTGGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAG 478
QY 704 CGGCTTAAGAGAGAGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 763
D 479 CGGCTTAAGAGAGAGAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 538
QY 764 GGTTCCTGACCTCTGACCT 783
D 539 GGTTCCTGACCTCTGACCT 558

RESULT 10
BM841897
LOCUS
DEFINITION BM841897 522 bp mRNA linear EST 06-MAR-2002
K-EST0119216 S13KMS5 Homo sapiens cDNA clone S13KMS5-34-A05 5',
mRNA sequence.
ACCESSION BM841897
VERSION BM841897.1 GI:19198306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 522)
JOURNAL Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
COMMENT Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
```


Db 165 GAGCGCACCCGCGCGCGCG-CCCCAGCCCCATGGACCTCCGAGCGGGGACTCGTGG 223
Qy 122 GGGATGTTAGCTTGTCTATGACAGGTGTGTGGACCTCCCTGACAGTCCAGCTCTTAAT 181
Db 224 GGGATGTTAGCTTGTCTATGACAGGTGTGTGGACCTCCCTGACAGTCCAGCTCTTAAT 283
Qy 182 GGCACAGAGATCCAGGCGCTGGCGCTCCATCCAGAAAACCTATGACCTCACCGCTAC 241
Db 284 GGCACAGGAGACCCAGGCGCTGGCGCTCCATCCAGAAAACCTATGACCTCACCGCTAC 343
Qy 242 CTGAGGATCAACCTCCGAGCTTGTGGACCTTGTGGACCTTGTGGACCTTGTGGACCTTGTTC 301
Db 344 CTGAGGATCAACCTCCGAGCTTGTGGACCTTGTGGACCTTGTGGACCTTGTGGACCTTGTTC 403
Qy 302 AACGAGCTGACCTTCAATCTCTGACTGGGGGAGAAAACCTTGGCCAGGCGCACCGTTC 361
Db 404 AACGAGCGAGACTTCAACCTTCCCGCTGGGGGAGAGACTTGGCCAGGCGCACCTGTT 463
Qy 362 AACTTGAAGTGTGGCGAGCTCAATGACAGGTGTGGCTGACCCAGAACTATGAGCG 421
Db 464 GACTTGGAGGTGTGGCGAGCTCAATGACAACTTGGCGCTGACCCAGAACTAGAGGCC 523
Qy 422 TACAGTCACTCTGTGTACTTGGTGGCTTCAACCTTCAACCTTCAAGCTG 467
Db 524 TACAGCACCTTCTGTGTACTTGGTGGCTTCAACCTTCAACCTTCAAGCTG 569

RESULT 15
BM363136
LOCUS 458 bp mRNA linear EST 10-JAN-2002
DEFINITION BS320054A10B08 Subtracted Lewin Cattle Spleen cDNA clone
BM363136
ACCESSION BM363136.1 GI:18106505
VERSION EST.
KEYWORDS cow.
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1. (bases 1 to 458)
Lewin,H.A., Soares,M.B., Pardinas,J., Liu,L. and Larson,J.H.
Subtracted Lewin Cattle Spleen ESTs
Unpublished (2002)
Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for Cattle EST sequencing was provided by the USDA National
Research Initiative, Project No. 98-35205-6644, and a grant from
the Japanese Ministry of Agriculture Fisheries and Forestry to
H.A.Lewin and J.E.Womack. Base-calling/Quality scores: PHRED form
Washington University Genome Center. Vector-trimming: Cross Match
from Washington University Genome Center PHRAP suite. This sequence
is vector free and at least 200bp in length.
Insert length: 458 Std Error: 0.00
Plate: BS320054A10 row: E column: 08
Seq primer: CGCAGCTCGAAATTAACCC
High quality sequence stop: 458.
Location/Qualifiers
1. .458
/organism="Bos taurus"
/strain="Angus"
/db_xref="taxon:9913"
/clone="BS320054A10E08"
/clone_lib="Subtracted Lewin Cattle Spleen"
/sex="Female"
/dev_stage="Adult"

FEATURES
source

/note="Organ: Spleen; Vector: pBluescript SK(+/-); Site 1:
EcoRI; Site 2: XhoI; This BS cDNA library was generated by
subtraction of the original non-normalized bovine spleen
library with 16,800 previously sequenced clones from a
bovine Placenta cDNA library. The original library was
constructed as described by Band et al (2000), Genome
Research 10(9): 1359-1368."
BASE COUNT 82 a 182 c 125 g 69 t
ORIGIN
Query Match 46.1%; Score 377.8; DB 13; Length 458;
Best Local Similarity 89.6%; Pred. No. 3.6e-79;
Matches 406; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
Qy 187 AGGAGATCCAGGCGCTGGCGCTCCATCCAGAAAACCTATGACCTCACCGCTACCTCGA 246
Db 6 AGGAGATCCAGGCGCTGGCGCTCCATCCAGAAAACCTATGACCTCACCGCTACCTCGA 65
Qy 247 GCATCAATCCGACAGCTTAGCTGGGAGCTTACCTGAACTACCTGGGGCCCCCTTTCAACGA 306
Db 66 GCACGAGCTCGGACAGCTTGGCTGGGAGCTTACCTGAACTACCTGGGGCCCCCTTTCAACGA 125
Qy 307 GCCTGACTTCAATCTCTCTGACTGGGGGAGAAAACCTTGGCCAGGCGCACGGTCAACTT 366
Db 126 GCCTGACTTCAATCTCTCTGACTGGGGGAGAAAACCTTGGCCAGGCGCACGGTCAACTT 185
Qy 367 GGAAGTGTGGCGAGAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGGCTACAG 426
Db 186 GGAGTGTGGCGAGAGCTTCAACGACAACTCGGGCTGACCCAGAACTATGAGGGCTACAG 245
Qy 427 TCACCTCTCTGTGTACTTGGCTGGCGCTCAACCTGAGCTGCCACAGCTGAACTCCGACG 486
Db 246 CCACCTCTCTGTGTACTTGGCTGGCGCTCAACCTGAGCTGCCACAGCTGAACTCCGACG 305
Qy 487 TAGCTGGCGCACTTCTGTACAGCTTCCAGGGCTGCTGGGCGAGCAATTGCAAGTGTCTAT 546
Db 306 CAGCTTGGCGCACTTCTGTACAGCTTCCAGGGCTGCTGGGCGAGCATCGGGCGTCTAT 365
Qy 547 GGCAGCTTGGCTTACCCACTTGGCGCTGAGCTTCCAGGGGAGCTGAGCGCTGAGCGCTG 606
Db 366 GGCAGCTTGGCTTACCCACTTGGCGCTGAGCTTCCAGGGGAGCTGAGCGCTGAGCGCTG 425
Qy 607 TGGCCCTGCGCCACAGTGAAGTCTTCTCCAGAAAGAT 639
Db 426 TGGCCCTGCGCCACAGTGAAGTCTTCTCCAGAAAGAT 458

Search completed: February 1, 2003, 12:59:17
Job time : 1669.54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 208.568 Seconds
(without alignments)
8843.084 Million cell updates/sec

Title: US-09-931-704-4
Perfect score: 819
Sequence: 1 tattattaagcttcgcgg.....aggccagtcagctgtgctt 819

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	819	100.0	819	19	AAV47512
2	819	100.0	819	19	AAV47512
3	819	100.0	819	21	AAV22654
4	819	100.0	819	24	AAV39483
5	674	82.3	881	22	AAH74484
6	674	82.3	1790	21	AAV88546
7	669.4	81.7	797	19	AAV47510
8	669.4	81.7	797	19	AAV22652
9	669.4	81.7	797	21	AAV39481

10	669.4	81.7	797	24	ABK11647	Human cDNA encodin
11	660	80.6	768	22	AAH99772	Human protein enco
12	646.4	78.9	648	21	AAA88547	Mouse interleukin-
13	627	76.6	1710	20	AAH16161	Human cardiotroph
14	600.2	73.3	729	22	AAO4201	Human cardiotroph
15	583.2	71.2	1008	22	AAK51548	Human polynucleoti
16	577	70.5	968	22	ABA09140	Human cardiotroph
17	577	70.5	968	22	ABA09140	Human polynucleoti
18	429.6	52.5	5087	21	AAK52532	Human polynucleoti
19	429.6	52.5	5087	24	AAH1648	Human NNT-1 DNA.
20	429.6	52.5	5088	19	AAV47511	Human novel neurot
21	429.6	52.5	5088	19	AAV22653	Human neurotrophic
22	419.8	51.3	495	22	ABA71773	Human genomic DNA
23	419.8	51.3	495	22	ABA71773	Human fetal liver
24	419.8	51.3	495	22	AAK20142	Probe #16321 for g
25	419.8	51.3	495	22	AAK46202	Human brain expres
26	419.8	51.3	495	22	AAI25564	Human bone marrow
27	419.8	51.3	495	22	AAI52108	Probe #15497 for g
28	419.8	51.3	495	24	ABS20513	Human genome-deriv
29	299.2	36.5	492	22	ABA59256	Human fetal liver
30	299.2	36.5	492	22	ABA27996	Probe #6462 for ge
31	299.2	36.5	492	22	AAK07469	Human brain expres
32	299.2	36.5	492	22	AAK33253	Human bone marrow
33	299.2	36.5	492	22	AAI16384	Probe #6317 for ge
34	299.2	36.5	492	22	AAI39047	Probe #7733 used t
35	299.2	36.5	492	24	ABS08085	Human genome-deriv
36	236	28.8	283	22	AAH23145	Osteoarthritis tis
37	65	7.9	65	24	ABN56698	Mouse spliced tran
38	64	7.8	1260	24	ABQ16902	Oligonucleotide fo
39	64	7.8	1260	24	ABQ16903	Oligonucleotide fo
40	52	6.0	396	20	AAK16162	Human cardiotroph
41	49.2	6.3	627	20	AAK60796	Human cardiotroph
42	47.4	5.8	627	20	AAK60797	Human DNAX interle
43	43.2	5.3	6803	18	AAH85473	Genomic hNET seque
44	43.2	5.3	6803	19	AAV16305	Genomic DNA encodi
45	42.4	5.2	1578	24	ABN96943	Gene #3441 used to

ALIGNMENTS

RESULT 1
AAV47512
ID AAV47512 standard; CDNA; 819 BP.
AC AAV47512;
XX
XX
DT 09-NOV-1998 (first entry)
DE Mouse neurotrophic factor NNT-1 CDNA.
XX

NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;
Alzheimer's disease; Parkinson's disease; Huntington's disease;
amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
peripheral neuropathy; dystrophy; neural retina degeneration;
common variable immunodeficiency; CVID; selective IGA deficiency;
hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
therapy; ss.

XX Mus sp.

XX Key Location/Qualifiers

FT CDS 95..772

FT sig_peptide /*tag= a

FT mat_peptide /*tag= b

XX /*tag= c

XX WO9833922-A1.

XX 06-AUG-1998.

XX

XX

XX	Key	Location/Qualifiers
FF	CDS	95..772
FT		/*tag= a
FT		/product= "NNT1"
PN		WO200215977-A2.
XX		
PD		28-FEB-2002.
XX		
PF		17-AUG-2001; 2001WO-US25906.
XX		
PR		18-AUG-2000; 2000US-226436P.
XX		
PR		16-AUG-2001; 2001US-0931704.
XX		
PA		(AMGE-) AMGEN INC.
XX		
PI		Senaldi G;
XX		
PI		WPI; 2002-280867/32.
DR		P-PSDB; AAU78177.
XX		
PT		Treating Immunoglobulin E-related disease, modulating IgE levels in a patient, preventing IgE-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient
XX		
PS		Claim 2; Fig 4; 63pp; English.
XX		
CC		The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor. The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type I allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular restenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, lupus, diabetes, inflammatory autoimmune disease, multiple sclerosis, and inflammatory joint disease, bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence encodes mouse NNT1.
XX		
SQ		Sequence 819 BP; 156 A; 288 C; 218 G; 157 T; 0 other;
	Query Match	100.0%; Score 819; DB 24; Length 819;
	Best Local Similarity	100.0%; Pred. No. 6e-196;
	Matches 819; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1	TATTATTAAGCTTGCCGGAGCGCGGGCTCGCCCTCCCACTCCGCCAGCCTCTGGAGA 60
Db	1	TATTATTAAGCTTGCCGGAGCGCGGGCTCGCCCTCCCACTCCGCCAGCCTCTGGAGA 60
QY	61	GGAGCGCGCCCGCGCGGCCCGCCCGCCAGCCCCATGGACCTCCGAGCAGGGACTCGT 120
Db	61	GGAGCGCGCGCCCGCGCGGCCCGCCCGCCAGCCCCATGGACCTCCGAGCAGGGACTCGT 120
QY	121	GGGGATGTTAGCTTGCTATGACCGTGCTGTGGCACTCCCTGCGAGTGCACGCTCTAA 180
Db	121	GGGGATGTTAGCTTGCTATGACCGTGCTGTGGCACTCCCTGCGAGTGCACGCTCTAA 180
QY	181	TCGCACAGGAGATCCAGGCCCTGGCCCTCCATCCAGAAAAACCTATGACCTCACCCGCTA 240

Db 181 TCACAGAGGATCCAGGCGCTGCGCCCTCCATCCAGAAACCTATGACCTCACCCGCTA 240
 QY 241 CTTGGAGCATCAACTCCGACGCTTAGCTGGGACCTACTGAACCTACTGGGGCCCCCTTT 300
 Db 241 CTTGGAGCATCAACTCCGACGCTTAGCTGGGACCTACTGAACCTACTGGGGCCCCCTTT 300
 QY 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGGGCCACGGT 360
 Db 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTGCCAGGGGCCACGGT 360
 QY 361 CAACCTGGAAAGTGGGAGGAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGG 420
 Db 361 CAACCTGGAAAGTGGGAGGAGCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGG 420
 QY 421 GTACAGTCACTCTGTTGTTACTTGGTGGCCCTCAACCGTCAAGCTGCCACAGCTGAAT 480
 Db 421 GTACAGTCACTCTGTTGTTACTTGGTGGCCCTCAACCGTCAAGCTGCCACAGCTGAAT 480
 QY 481 CCGAGCTAGCTGCGCCACTTCTGTACACGCTCCAGGCTCGTGGGAGCATTCGAGG 540
 Db 481 CCGAGCTAGCTGCGCCACTTCTGTACACGCTCCAGGCTCGTGGGAGCATTCGAGG 540
 QY 541 TGTATGGGAGCTTGGCTTACCACTGCCCCAGGCTCTGCCAGGAGTCAAGCAAGAT 600
 Db 541 TGTATGGGAGCTTGGCTTACCACTGCCCCAGGCTCTGCCAGGAGTCAAGCAAGAT 600
 QY 601 GGGCCCTGGCCCTGCCACAGTACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAA 660
 Db 601 GGGCCCTGGCCCTGCCACAGTACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAA 660
 QY 661 GGAGCTGCAGACTGGCTATGGCTTCCAGGAGGACTTCAACCGCTTAAGAAAGAT 720
 Db 661 GGAGCTGCAGACTGGCTATGGCTTCCAGGAGGACTTCAACCGCTTAAGAAAGAT 720
 QY 721 GCAGCTCCAGAGCTTCACTCAGCTTGGAGGACATGTTCTGACCTCTGAC 780
 Db 721 GCAGCTCCAGAGCTTCACTCAGCTTGGAGGACATGTTCTGACCTCTGAC 780
 QY 781 CTTTAAACCCACACTCCAGGCCAGTCAAGTGTGCTT 819
 Db 781 CTTTAAACCCACACTCCAGGCCAGTCAAGTGTGCTT 819

RESULT 5

AAH74484

ID AAH74484 standard; DNA; 881 BP.

AC AAH74484;

XX 15-OCT-2001 (first entry)

XX Nucleotide sequence of a human NNT-1 protein.

KW NNT-1; CLF-1; SCNTFRalpha; nervous system; neuron; nervous system;
 KW neuro-muscular function; tumour; immune system; haematopoietic system;
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;
 KW blastocyst implantation; thrombosis; retinal disease;
 KW retinal pigmentosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 174..851

XX /*tag= a

XX /product= "NNT-1"

XX WO200155172-A2.

XX 02-AUG-2001.

XX

XX

XX

XX

PF 26-JAN-2001; 2001WO-FR00253.
 XX 27-JAN-2000; 2000FR-0001035.
 PR 12-OCT-2000; 2000FR-0013089.
 XX (FABR) FABRE MEDICAMENT SA PIERRE.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;
 PI WPI; 2001-488773/53.
 XX P-PSDB; AAG63543.
 DR A complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha
 XX protein useful to treat neurodegenerative disease including Parkinson's
 PT and Huntington's, obesity and cancer -
 PT Disclosure; Page 57-58; 67pp; French.
 XX The present sequence encodes a human NNT-1 protein. The specification
 CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or
 CC SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate
 CC activity of the SCNTFRalpha/gp130/LiFRbeta receptor complex, or to
 CC induce phosphorylation of the tyrosine of gp130 and LiFRbeta.
 CC particularly where cells expressing the receptor complex are in the
 CC central or peripheral nervous system, in neurons implicated in
 CC neuro-muscular function or in skeletal muscle. The complex or
 CC antibodies are also used to decrease the survival, growth or
 CC proliferation of tumour cells or to facilitate the proliferation and/or
 CC inhibit differentiation of cells stocks. The complex is also used to
 CC modulate activity of the gp130/LiFRbeta receptor or cells expressing
 CC that receptor, particularly those cells implicated in the immune,
 CC haematopoietic, nervous or reproductive system, the liver or skeletal
 CC muscle. Molecules of the invention may be used to prevent or treat
 CC neurodegenerative diseases including amyotrophic lateral sclerosis,
 CC Parkinson's and Huntington's disease, to repair or regenerate nervous
 CC or muscular tissue or to maintain muscular mass in paralysis patients.
 CC They may also be used to treat cancer, obesity and associated diseases,
 CC and to improve fertility, particularly to avoid endometriosis and/or
 CC assist blastocyst implantation, thrombosis, or retinal disease,
 CC particular retinal pigmentosis.

SQ Sequence 881 BP; 158 A; 318 C; 246 G; 159 T; 0 other;
 Query Match 82.3%; Score 674; DB 22; Length 881;
 Best Local Similarity 92.2%; Pred. No. 1.5e-159;
 Matches 721; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

QY 2 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGGACTCTGGAGAG 61
 Db 82 ATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGGACTCTGGAGAG 141

QY 62 GAGCGGCG 121
 Db 142 GAGCGGCG 200

QY 122 GGGATGTTAGCTTGCCTATGACGCTGCTGCGGAGCTTCCCTCGAGTCCAGCTCTTAAT 181
 Db 201 GGGATGTTAGCTTGCCTATGACGCTGCTGCGGAGCTTCCCTCGAGTCCAGCTCTTAAT 260

QY 182 CGCAGAGATCCAGGCG 241
 Db 261 CGCAGAGATCCAGGCG 320

QY 242 CTGGAGCATCACTCCGAGCTTAGCTGGGACCTACTGAACTACCTGGGGCCCCCTTTC 301
 Db 321 CTGGAGCATCACTCCGAGCTTAGCTGGGACCTACTGAACTACCTGGGGCCCCCTTTC 380

QY 302 AACGAGCTTGAATTCCTCTCGACTGGGGGAGAACTCTGCGGAGGAGGAGGAGGAGG 361
 Db 381 AACGAGCTTGAATTCCTCTCGACTGGGGGAGAACTCTGCGGAGGAGGAGGAGGAGG 440

QY 362 AACTTGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 421

Sun Feb 2 08:32:21 2003

us-09-931-704-4.rng

Db 441 GACTTGGAGGTGTGGCGAAGCCTCAATGACAAACTCGGGCTGACCCAGAACTACGAGGCC 500
Qy 422 TACAGTCACCTCTGTGTTACTTGGCTGCTCAACCGTCAGGCTGCCAGCTGAATC 481
Db 501 TACAGCCACCTTCTGTGTTACTTGGCTGCTCAACCGTCAGGCTGCCAGCTGAATC 560
Qy 482 CGACGTAGCTGGCCACCTTCTGTACAGCCTCCAGGCTGCTGGGCGACATTCAGGT 541
Db 561 CGCGCAGCTGGCCACCTTCTGTACAGCCTCCAGGCTGCTGGGCGACATTCAGGT 620
Qy 542 GTATGGGAGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 601
Db 621 GTATGGGAGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 680
Qy 602 GCGCTGGGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 661
Db 681 ACTCTGGGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 740
Qy 662 GAGCTGCAGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 721
Db 741 GAGCTGCAGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 800
Qy 722 CAGCTTCAGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 781
Db 801 CAGCTTCAGCTGGCTTACCTACCTGCTCCAGGCTGCTGGGCGACATTCAGGT 860
Qy 782 CT 783
Db 861 TT 862
RESULT 6
ID AAA88546 standard; DNA; 1790 BP.
XX AAA88546;
AC AAA88546;
XX 22-JAN-2001 (first entry)
DT Human interleukin-B60 (IL-B60) gene.
XX Interleukin-B60; IL-B60; human; cytokine; chromosome 11;
XX cytokine-like factor-1; haematopoietic; inflammation;
XX antiinflammatory; autoimmune disease; therapy; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
PH 162..809
FT CDS /tag= a
FT sig_peptide 162..212
FT /tag= b
FT mat_peptide 213..806
FT /tag= c
XX WO200053631-A1.
XX 14-SEP-2000.
XX 09-MAR-2000; 2000WO-US06182.
XX 11-MAR-1999; 99US-0267901.
XX (SCHE) SCHERING CORP.
XX Oppmann B, Timans JC, Kastelein RA, Bazan JF;
XX WPI; 2000-587426/55.
XX P-PSDB; AAB19686.
XX Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,
XX polypeptides, and nucleic acids, useful in research, diagnosis and for

PT treating inflammatory and autoimmune disorders -
XX Claim 17; Page 15-16; 97pp; English.
XX The present sequence is that of DNA encoding human interleukin-B60
CC (IL-B60, see AAB19586), a novel, small soluble cytokine-like protein
CC of 198 amino acids that exhibits structural motifs characteristic
CC of a member of the long-chain cytokines, and which shows homology
CC to granulocyte colony stimulating factor and interleukin-6. IL-B60
CC may have either stimulatory or inhibitory effects on haematopoietic
CC cells, including e.g. lymphoid cells, such as T-cells, B-cells,
CC natural killer cells, macrophages, dendritic cells, haematopoietic
CC progenitors, etc. Methods are provided for modulating the
CC physiology or development of a cell or tissue culture cells by
CC contacting the cell with an agonist or antagonist of IL-B60 or an
CC agonist of antagonist of a complex of mature IL-B60 and its
CC partner, cytokine-like factor-1 (CLF-1, see AAB19588). The
CC IL-B60/CLF-1 cytokine serves as a key physiological factor in motor
CC neuron development and regeneration. IL-B60, its agonists and
CC antagonists may be used to treat inflammatory or autoimmune
CC disorders and also for drug screening. The IL60B gene maps to
CC human chromosome 11.
XX Sequence 1790 BP; 381 A; 560 C; 474 G; 375 T; 0 other;
SQ Query Match 82.3%; Score 674; DB 21; Length 1790;
Best Local Similarity 92.2%; Pred. No. 1.8e-159;
Matches 721; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
Qy 2 ATTATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGAGGAGGAGGAG 61
Db 40 ATTATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGAGGAGGAGGAG 99
Qy 62 GAGCG 121
Db 100 GAGCG 158
Qy 122 GGGATTTAGCTTGCCTATGCGAGGTGCTGTGGCACTCCCTCAGTGCCAGCTTAAAT 181
Db 159 GGGATTTAGCTTGCCTATGCGAGGTGCTGTGGCACTCCCTCAGTGCCAGCTTAAAT 218
Qy 182 CGCAGGAGATCCAGGCGCTGCGCGCTCCATCCAGAAACCTATGACCTCACCGCTAC 241
Db 219 CGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 278
Qy 242 CTGGAGCATCACTCCGCGAGCTTAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
Db 279 CTGGAGCATCACTCCGCGAGCTTAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 338
Qy 302 AACGAGCTGAGCTTCAATCTCTCGAGCTGGGGGAGAGAACTCTGCCAGGGCCACGGTC 361
Db 339 AACGAGCTGAGCTTCAACCTCTCGAGCTGGGGGAGAGAGACTCTGCCAGGGCCACGGTC 398
Qy 362 AACTTGGAGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 421
Db 399 GACTTGGAGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 458
Qy 422 TACAGTCACCTCTGTGTTACTTGGCTGCTCCAGGCTGCTGGGCGACATTCAGGT 481
Db 459 TACAGTCACCTCTGTGTTACTTGGCTGCTCCAGGCTGCTGGGCGACATTCAGGT 518
Qy 482 CGAGGTAGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 541
Db 519 CGCGCAGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 578
Qy 542 GTCATGGCAGCTTGGGCTACCCACTGCGCGCTGCGCTGGGAGTGAACCACTTGG 601
Db 579 GTCATGGCAGCTTGGGCTACCCACTGCGCGCTGCGCTGGGAGTGAACCACTTGG 638
Qy 602 GCGCTGGGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 661
Db 639 ACTCTGGGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 698


```
RESULT 8
AAV22652
ID AAV22652 standard; cDNA; 797 BP.
XX
AC AAV22652;
XX
DT 13-JUL-1998 (first entry)
XX
XX cDNA encoding human neurotrophic factor NNT-1.
XX
XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 90..767
FT sig_peptide /*tag= a
FT mat_peptide /*tag= b
FT /*tag= c
XX
XX US57411772-A.
PN
XX
XX 21-APR-1998.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX (AMGE-) AMGEN INC.
XX
XX Chang M;
XX
XX WPI; 1998-260526/23.
DR P-PSDB; AAV56141.
XX
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
PT useful for stimulating growth of motor and sympathetic neurons
XX
XX Disclosure; Fig 1; 41pp; English.
XX
XX The present sequence encodes a human neurotrophic factor, designated
CC NNT-1, which is capable of stimulating growth of motor or sympathetic
CC neurons. The NNT-1 protein is useful in the treatment of neurological
CC diseases characterised by the degeneration and death of particular
CC classes of neurons. These diseases specifically include Parkinson's
CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
CC stroke and various degenerative disorders affecting vision.
XX
XX Sequence 797 BP; 139 A; 297 C; 218 G; 143 T; 0 other;
SQ
Query Match 81.7%; Score 669.4; DB 19; Length 797;
Best Local Similarity 92.0%; Pred. No. 2.1e-158;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
5 ATTTAAAGCTTCGGAGCGCGGCTCGCCCTCCCACTCGGCAGCCTCTGGAGAGGAG 64
1 ATTTAAAGCTTCGGAGCGCGGCTCGCCCTCCCACTCGGCAGCCTCTGGAGAGGAG 60
65 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
61 CGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
125 ATGTTAGCTTGCTATGACGCGTGTGGACCTCCCTGCACTGCGAGCTCTTAATCGC 184
120, ATGTTAGCTGCTGTGACGCGTGTGGACCTCCCTGCACTGCGAGCTCTCAATCGC 179
185 ACAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 244
180 ACAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
```

```
QY 245 GAGATCAACTCCGACCTTAGCTGGACCTACCTGAACCTACCTGGGGCCCCCTTTCAAC 304
DB 240 GAGACCAACTCCGACCTTAGCTGGACCTATCTGAACCTACCTGGGGCCCCCTTTCAAC 299
QY 305 GAGCCTGACTTCAATCTCTCGACTGGGGCAGAACTCTGCCAGGGCCACGGTCAAC 364
DB 300 GAGCCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTGTTGAC 359
QY 365 TTGGAAGTGTGGCGAAGCCTCAATGACAGGCTGGGGCTGAGCCAGAACTATGAGGCGTAC 424
DB 360 TTGAGAGTGTGGCGAAGCCTCAATGACAACTGGGGCTGAGCCAGAACTATGAGGCGTAC 419
QY 425 AGTCACCTCTCTGTGTTACTTGGCTGGCTCAACCGTCAAGGCTGCCAGCTGAATCCGA 484
DB 420 AGCCACCTCTCTGTGTTACTTGGCTGGCTCAACCGTCAAGGCTGCCAGCTGCCGCTGCGC 479
QY 485 CGTAGCCTGGCCCACTTCTGTACAGAGCCTCCAGGGCCCTGCTGGCAGCATTTGAGGCTGTC 544
DB 480 CGCAGCCTGGCCCACTTCTGTACAGAGCCTCCAGGGCCCTGCTGGCAGCATTTGAGGCGTTC 539
QY 545 ATGGCGAGCGCTTGGCTACCCACTGCCAGCCTCTGCCAGGGAGCTGAGCCAGCTGGGCC 604
DB 540 ATGGCAGCTCTGGGCTACCCACTGCCAGCCTCTGCCAGGGAGCTTCTGGGCTGCTGAAGGAG 599
QY 605 CCTGGCCCTGGCCCACTGAGTACTTCTCCAGAAAGATGAGTACTTCTGGCTGCTGAAGGAG 664
DB 600 CCTGGCCCTGGCCCACTGAGTACTTCTCCAGAAAGATGAGTACTTCTGGGCTGCTGAAGGAG 659
QY 665 CTGCAGACTGGCTATGGGTTTCAGCCAGGAGCTTCAACCGGCTTAAGAAAGAGATGCGAG 724
DB 660 CTGCAGACTGGCTATGGGTTTCAGCCAGGAGCTTCAACCGGCTTAAGAAAGAGATGCGAG 719
QY 725 CCTCCAGCAGCTTTCAGTCACTCCCTGCACTTGGAGGCGCATGGTTTCTGACCTCTGACCTT 783
DB 720 CCTCCAGCAGCTTTCAGTCACTCCCTGCACTTGGAGGCGCTCATGGGCTTCTGACTTCTGACCTT 778
RESULT 9
AAA39481
ID AAA39481 standard; cDNA; 797 BP.
XX
AC AAA39481;
XX
DT 24-AUG-2000 (first entry)
XX
XX Human NNT-1 cDNA.
XX
XX NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
KW retinopathy; immune disorder; hematopoietic disorder; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 90..767
FT /*tag= a
FT /*product= "NNT-1"
XX
XX US6054294-A.
XX
XX 25-APR-2000.
XX
XX 12-DEC-1997; 97US-0988819.
XX
XX 03-FEB-1997; 97US-0792019.
XX
XX (AMGE-) AMGEN INC.
XX
XX Chang M;
```


occurring variant, and diagnosing an IGE-related disease or
 susceptibility of an IGE-related disease based on the presence or amount
 of expression of the polypeptide and a pharmaceutical composition for use
 in treating IGE-related disease, comprising the NNT1 inhibitor.
 The NNT1 inhibitor is useful for preventing and treating IGE-related
 disease, modulating IGE levels, and treating allergic diseases e.g.
 type 1 allergic disease, allergic rhinitis, eczema, dermatitis,
 pollinosis, asthma, immune diseases and disorders, diseases involving
 abnormal cell proliferation including cancer, arteriosclerosis and
 vascular restenosis, diseases and conditions relating to dysfunction of
 immune system including rheumatoid arthritis, psoriatic arthritis,
 inflammatory arthritis, osteoarthritis, inflammatory joint disease,
 autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
 bowel disease, transplant rejection, and graft versus host disease, and
 reproductive diseases and disorders including infertility, miscarriage,
 preterm labour and delivery, and endometriosis. The present sequence
 encodes human NNT1.

Seq	Sequence	797 BP	139 A	297 C	218 G	143 T	0 other
Query Match	81.7%	Score	669.4	DB	24	Length	797
Best Local Similarity	92.0%	Pred. No.	2.1e-158				
Matches	717	Conservative	0	Mismatches	61	Indels	1
Gaps							
QY	5	ATTTAAAGCTTTCGCGAGCGCGGCTCGCCCTCCACCTCCGACGACCTCTTGGGAGGAG	64				
DB	1	ATTTAAAGCTTTCGCGAGCGCGGCTCGCCCTCCACCTCCGACGACCTCTTGGGAGGAG	60				
QY	65	CGCGCGCGCGCGCGCGCGCGCGCGCCACATGAGCCCTCCGAGCAGGGAGCTCTGTGGGGG	124				
DB	61	CGCGACCGCGCGCGCGCGCGCGCGCGCGCCACATGAGCCCTCCGAGCAGGGAGCTCTGTGGGGG	119				
QY	125	ATGTAGCTTGGCTATGACAGGTGTGTGGGACCTCCCTGCAGTGCCAGCTCTTAATCGC	184				
DB	120	ATGTAGCTGTGCTGTGACAGGTGTGTGGGACCTCCCTGCAGTGCCAGCTCTTAATCGC	179				
QY	185	ACAGGAGATCCAGGCGCTCGGCCCTCCATCCAGAAACCTATGACCTCACCGGCTACCTG	244				
DB	180	ACAGGGGACCGAGGGCTGGCCCTCATTCAGAAACCTATGACCTCACCGGCTACCTG	239				
QY	245	GAGCATCAACTCCGACGCTTAGCTGGGACCTAAGCTCTCTGGGGGCCCCCTTTCAAC	304				
DB	240	GAGCACAACCTCCGACGCTTGGCTGGGACCTAAGCTCTCTGGGGGCCCCCTTTCAAC	299				
QY	305	GAGCCTGACTTCAATCCTCTCGACTGGGGGACAGAACTCTGCCAGGGCCACGGTCAAC	364				
DB	300	GAGCCAGACTTCAACCCCTCCCGGCTGGGGGACAGACTCTGCCAGGGCCACGGTCAAC	359				
QY	365	TTGCGAAGTGGCGAAGCCTCAATGACAGGCTGCGGTGACCCAGAACTATGAGGCGTAC	424				
DB	360	TTGCGAAGTGGCGAAGCCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCGTAC	419				
QY	425	AGTCACCTCTGTGTATTGTGGTGGGCTCAACCGTCAAGCTGGCAGCTGAATCCGA	484				
DB	420	AGCCACCTTCTGTGTATTGTGGTGGGCTCAACCGTCAAGCTGGCAGCTGAATCCGA	479				
QY	485	CGTAGCTGGCCACCTTCTGTACAGGCTCCAGGCGCTGTGGGACGACTTCAGAGTGTC	544				
DB	480	CGAGCTTGGCCACCTTCTGTACAGGCTCCAGGCGCTGTGGGACGACTTCAGAGTGTC	539				
QY	545	ATGGCAGCTTGGCTACCCACTGCCCGAGCCTCTGCAGGGACTGAGCCAGCCTGGGCC	604				
DB	540	ATGGCAGCTTGGCTACCCACTGCCCGAGCCTCTGCAGGGACTGAGCCAGCCTGGGAC	599				
QY	605	CTTGGCCCTGCCACAGTGACTTCTCCAGAGATGATGACTTCTGGCTGTCAAGGAG	664				
DB	600	CTTGGCCCTGCCACAGTGACTTCTCCAGAGATGATGACTTCTGGCTGTCAAGGAG	659				
QY	665	CTGCAGACCTGGCTATGCGGTTACGCCAAGGACTTCAACCCGGCTTAAGAAAGATGCG	724				
DB	660	CTGCAGACCTGGCTATGCGGTTACGCCAAGGACTTCAACCCGGCTTAAGAAAGATGCG	719				
QY	725	CCTCCAGCAGCTTCACTGACCCCTGCACCTGTGGAGGACATGGTTTCTGACCTCTGACCT	783				

720 CCTCCAGCAGTGCAGTCAACCTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTT 778

Db

RESULT 11
AAH99772
ID AAH99772 standard; cDNA, 768 BP.
XX AC
XX AC AAH99772;
XX DT 16-OCT-2001 (first entry)
XX DE
XX DE Human protein encoding cDNA sequence SEQ ID NO:607.
XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX KW antibacterial; endocrine; cardiant; central nervous system; virucide;
XX KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
XX KW antiagregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
XX KW dermatological; anti allergic; antiasthmatic; antidiabetic; cytostatic;
XX KW neuroprotective; antidepressant; nootropic; vaccine; inflammation;
XX KW immunostimulant; gene therapy; antisense therapy; antiparkinsonian; infection;
XX KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;
XX KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX KW allergic rhinitis; diabetes; multiple sclerosis; depression;
XX KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX KW neurological disorder; ss.
XX OS
XX OS Homo sapiens.
XX PN WO200153455-A2.
XX PD
XX PD 26-JUL-2001.
XX PF 22-DEC-2000; 2000WO-US35017.
XX XX
XX PR 23-DEC-1999; 99US-0471275.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PA (HYSE-) HYSEQ INC.
XX PI
XX PI Tang YT, Liu C, Drmanac RT;
XX XX
XX XX WPI; 2001-457603/49.
XX XX P-PSDB; AAM25831.
XX PT Isolated human polynucleotides encoding polypeptides, useful for the
XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX PS
XX PS Claim 1; Page 638; 1217pp; English.
XX XX
XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX CC AAH25963. The proteins can have activities based on the tissues and
XX CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
XX CC central nervous system; virucide; anti-HIV; fungicide; antitumor;
XX CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
XX CC antiulcer; osteopathic; dermatological; anti allergic; antiasthmatic;
XX CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine
XX CC production, the proteins and polynucleotides are useful for screening for
XX CC agonists or antagonists of a protein and for the treatment and diagnosis
XX CC of disorders associated with the activity of a protein e.g. inflammatory
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX CC infections, autoimmune, genetic diseases, haematopoietic disorders,
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, CC

[illegible]

Sun Feb 2 08:32:21 2003

20-OCT-2000; 2000US-06933325.
 30-NOV-2000; 2000US-0728422.
 (HYSE-) HYSEQ INC.
 Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 Zhao Q, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 WPI; 2001-476283/51.
 P-PSDB; AM78415.
 Nucleic acids encoding polypeptides with cytokine-like activities,
 useful in diagnosis and gene therapy -
 Claim 1; Page 711-712; 6221pp; English.
 The invention relates to polynucleotides (AAK51456-AAK53435) and the
 encoded polypeptides (AM78323-AM80302) that exhibit activity elating to
 cytokine, cell proliferation or cell differentiation or which may induce
 production of other cytokines in other cell populations. The
 polynucleotides and polypeptides are useful in gene therapy, vaccines or
 peptide therapy. The polypeptides have various cytokine-like activities,
 e.g. stem cell growth factor activity, haematopoiesis regulating
 activity, tissue growth factor activity, immunomodulatory activity and
 activin/inhibin activity and may be useful in the diagnosis and/or
 treatment of cancer, leukaemia, nervous system disorders, arthritis and
 inflammation.
 Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 (AM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1008 BP; 183 A; 374 C; 247 G; 204 T; 0 other;

Query Match 71.2%; Score 583.2; DB 22; Length 1008;
 Best Local Similarity 90.8%; Pred. No. 9.5e-137;
 Matches 621; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 100 CCTCCGAGCAGGGACTCGTGGGGGATGTTAGCTTGCCTATGACAGGCTGTGGCACT 159
 DB 296 CCTGCCACAGGGGACTCGTGGGGGATGTTAGCGTCTGTGCAGCGTCTTGGCACT 355
 QY 160 CCTTCAGTGCAGCTCTTAATCGACAGAGATCCAGGCCCTGGCCCCCTCCATCCAGAA 219
 DB 356 CCTTCAGTGCAGCTCTTAATCGACAGAGATCCAGGCCCTGGCCCCCTCCATCCAGAA 415
 QY 220 AACCTATGACCTACCCGCTACCTGGAGGATCACTCCGAGCTTAGCTGGGACCTACT 279
 DB 416 AACCTATGACCTACCCGCTACCTGGAGGATCACTCCGAGCTTAGCTGGGACCTACT 475
 QY 280 GAACCTACCTGGGGCCCTTTTCAAGAGCTGACTTCAATCTCTCGACTGGGGGCGAGA 339
 DB 476 GAACCTACCTGGGGCCCTTTTCAAGAGCTGACTTCAATCTCTCGACTGGGGGCGAGA 535
 QY 340 AACTCTGCCCCAGGGCCAGCGTCAACTTGAAGTGTGGGAGAGCTCAATGACAGGCTGG 399
 DB 536 GACTCTGCCAGGGCCACTGTTGACTTGGAGGTGGCGAGGCTCAATGACAAACTGG 595
 QY 400 GCTGACCCAGAACTATGAGCGGTACAGTCACTCCCTGCTGTTACTTGGTGGCTCAACCG 459
 DB 596 GCTGACCCAGAACTATGAGCGGTACAGTCACTCCCTGCTGTTACTTGGTGGCTCAACCG 655
 QY 460 TCAGGCTGCCAGCTGAACCTCCGAGCTAGCTGGCCCACTTCTGTACCAGCTCCAGGG 519
 DB 656 TCAGGCTGCCAGCTGAACCTCCGAGCTAGCTGGCCCACTTCTGTACCAGCTCCAGGG 715
 QY 520 CCTGTGGGCGAGCTATGAGGTGTATGGGAGCGCTTGGCTAGCAGCTGCCCGGCTCT 579
 DB 716 CCTGTGGGCGAGCTATGAGGTGTATGGGAGCGCTTGGCTAGCAGCTGCCCGGCTCT 775
 QY 580 GCCAGGGAGCTAGCGAGCTGGGGCCCTGGCCCTGCCCACTGAGTCTTCTCCAGAGAT 639
 DB 776 GCTTGGGAGCTAGCGAGCTGGGGCCCTGGCCCTGCCCACTGAGTCTTCTCCAGAGAT 835

QY 640 GGATGACTTCTGCTGCTGAAGGAGCTGCAGACCTGGCTATGGCTTCAGCCAGGACTT 699
 DB 836 GGACGACTTCTGCTGCTGAAGGAGCTGCAGACCTGGCTATGGCTTCAGCCAGGACTT 895
 QY 700 CAACCCGCTTAAAGAAAGATGCGAGCCTCCAGCAGCTTCACTACCCCTGCACCTTGGAGGC 759
 DB 896 CAACCCGCTCAAGAAAGATGCGAGCCTCCAGCAGCTTCACTACCCCTGCACCTTGGAGGC 955
 QY 760 ACATGGTTTCTGACCTCTGACCT 783
 DB 956 TCATGGCTTCTGACTTCTGACCTT 979

Search completed: February 1, 2003, 09:12:12
 Job time : 211.568 secs


```
FILE REFERENCE: A-695
CURRENT APPLICATION NUMBER: US/09/931.704
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: US 60/226,436
PRIOR FILING DATE: 2000-08-18
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent in version 3.1
SEQ ID NO 3
LENGTH: 5087
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (137)..(138)
OTHER INFORMATION: product = "INTERVENING UNSEQUENCED REGION OF >1KB"
US-09-931-704-3

Query Match          52.5%; Score 429.6; DB 10; Length 5087;
Best Local Similarity 90.4%; Pred. No. 5.5e-105;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCTCCTCGACTGGGG 335
Db 3363 AGCTGAACCTACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTTCCCGCCTGGGG 3422
QY 336 CAGAAACTCTGCCAGGCCCCAGGCTCAACTTGAAGTGTGCGAAGCCTCAATGACAGGC 395
Db 3423 CAGAGACTCTGCCAGGCCCCACTTGTGACTTGGAGGTGTGCGAAGCCTCAATGACAAAC 3482
QY 396 TGGGGCTGACCCAGAACATGAGCGGTACAGTCACTCCTCTGTATTACTTGGTGGCTTCA 455
Db 3483 TGGGGCTGACCCAGAACATGAGCGGTACAGTCACTCCTCTGTATTACTTGGTGGCTTCA 3542
QY 456 ACCGTGAGGCTGCCACAGCTGAACTCCGAGCTAGCTGGCCCACTTGTATACCAAGCTCC 515
Db 3543 ACCGTGAGGCTGCCACAGCTGAGTGTGGCCCACTTGTATACCAAGCTCC 3602
QY 516 AGGGCTCTGCTGGGAGCATTTGAGTGTGATGGGAGCGCTTGGTACCCACTGCCCCAGC 575
Db 3603 AGGGCTCTGCTGGGAGCATTTGCGGGCGTGTGCGAGCTCTGGGTACCCACTGCCCCAGC 3662
QY 576 CTCTGCAGGAGTACGAGCAGCTGGGGCCCTGGCCCTGCCAGTGCAGTACTTCTTCAGA 635
Db 3663 CGCTGCTGGGAGTGAACCCACTTGGACTCTCGCCCTGCCAGTGCAGTACTTCTTCAGA 3722
QY 636 AGATGGATGACTTCTGGCTGTCTGAAGAGCTGCAGAGCTGCTATGGCTTGCAGCAAGG 695
Db 3723 AGATGGAGACTTCTGGCTGTCTGAAGAGCTGCAGAGCTGCTATGGCTTGCAGCAAGG 3782
QY 696 ACTTCAACCGGCTTAAAGAAGATGACAGCCTCCAGCAGCTTTCAGTCAACCTGCACTTGG 755
Db 3783 ACTTCAACCGGCTCAAGAAGAAGATGACAGCCTCCAGCAGCTGCACTCAACCTGCACTGG 3842
QY 756 AGGCACATGTTTCTGACCTCTGACCT 783
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

RESULT 4
US-09-864-761-23175/c
Sequence 23175, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aesomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
```

```
FILE REFERENCE: A-695
CURRENT APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 4917
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 23175
LENGTH: 495
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005849.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
OTHER INFORMATION: SWISSPROT HIT: O63086, EVALUATION 8.00e-03
OTHER INFORMATION: NT HIT: g11439486, EVALUATION 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AL752561.1, EVALUATION 0.00e+00
US-09-864-761-23175

Query Match          51.3%; Score 419.8; DB 10; Length 495;
Best Local Similarity 90.5%; Pred. No. 1e-102;
Matches 448; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 278 CTGAATACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCCTCTCGACTGGGGCA 337
Db 495 CTGAATACCTGGGGCCCCCTTTCAACGAGCAGACTTCAACCTTCCCGCCTGGGGCA 436
QY 338 GAAACTCTGCCAGGCCCCAGTCACTTGAAGTGTGCGAAGCCTCAATGACAGGCTG 397
Db 435 AAGACTCTGCCAGGCCCCACTTGTGACTTGGAGTGTGCGAAGCCTCAATGACAACTG 376
QY 398 CGGCTGACCCAGAACTATGAGCGCTCAGACTCCTCTCTGTACTTTCGCTGGCTCAAC 457
Db 375 CGGCTGACCCAGAACTATGAGCGCTCAGACTCCTCTCTGTACTTTCGCTGGCTCAAC 316
QY 458 CGTCAGGCTGCACAGCTGAACCTCCGAGCTAGCTGGCCCACTTCTGTACCAAGCTCCAG 517
```

Db 315 CGTCAGGCTGCCACTGCTAGCTGGCGCGAGCTGGCCCACTTCTGCACGAGCTCCAG 256
QY 518 GGCCTGCTGGGAGCATTCAGGTGTCATGGCAGCTTGGCTACCCAGCTCCAGCCT 577
Db 255 GGCCTGCTGGGAGCATTCAGGTGTCATGGCAGCTTGGCTACCCAGCTCCAGCCT 196
QY 578 CTGCCAGGACTGAGCAGCTGGGCGCCCTGGCCCTGCGCCACAGTACTTCTCCAGAAG 637
Db 195 CTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCGCCACAGTACTTCTCCAGAAG 136
QY 638 ATGATGACTTCTGGCTGTGAAGAGCTGCAGACTCTGGCTGTATGGCTTCAGCCAAAGGAC 697
Db 135 ATGAGCACTTCTGGCTGTGAAGAGCTGCAGACTCTGGCTGTATGGCTTCAGCCAAAGGAC 76
QY 698 TTCAACCGGCTTAAGAGAGATGCAGCTCCAGCAGCTTCAGTCACTCCCTGCCTGGAG 757
Db 75 TTCAACCGGCTTAAGAGAGATGCAGCTCCAGCAGCTTCAGTCACTCCCTGCCTGGAG 16
QY 758 GCACATGGTTCTGA 772
Db 15 GTCATGGCTTCTGA 1

RESULT 5
US-09-864-761-6462/c
; Sequence 6462, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6462
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
US-09-864-761-6462

Query Match 36.5%; Score 299.2; DB 10; Length 492;
Best Local Similarity 90.6%; Pred. No. 1.3e-70;
Matches 319; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 432 TCCTGTGTTACTTGGTGGCCTCAACCGTCAAGGCTGCCAGCTGAACCTCCAGCGTAGCC 491
Db 492 TCCTGTGTTACTTGGTGGCCTCAACCGTCAAGGCTGCCAGCTGAACCTCCAGCGTAGCC 433
QY 492 TGGCCCACTTCTGTACCAAGCCTCCAGGCGCTGCTGGGAGCAGATTGACGTGTCTATGGCGA 551
Db 432 TGGCCCACTTCTGTACCAAGCCTCCAGGCGCTGCTGGGAGCAGATTGACGTGTCTATGGCGA 373
QY 552 CGCTTGGCTACCCACTGCCCGCAGCCTCTGCCAGGAGCTGAGCCAGCTGGGCGCCCTGGGCC 611
Db 372 CTCTGGGCTACCCACTGCCCGCAGCCTGCTGGGAGCAGATTGACGTGTCTATGGCGC 313
QY 612 CTGCCCCAGTGACTTCTCCAGAAAGATGGATGACTTCTGGCTGTGAAGAGAGCTGCAGA 671
Db 312 CTGCCCCAGTGACTTCTCCAGAAAGATGGAGACTTCTGGCTGTGAAGAGAGCTGCAGA 253
QY 672 CTTGGCTATGGCGTTCAGCCCAAGGACTTCAACCGGCTTAAGAGAGAGATGACGCTCCAG 731
Db 252 CTTGGCTATGGCGTTCAGCCCAAGGACTTCAACCGGCTTAAGAGAGAGATGACGCTCCAG 193
QY 732 CAGCTTCAGTCACCTGCAGCTTGAGGAGCAGATGTTCTGACCTCTGACCT 783
Db 192 CAGCTTCAGTCACCTGCAGCTTGAGGAGCAGATGTTCTGACCTCTGACCT 141

RESULT 6
US-09-765-231A-75
; Sequence 75, Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
; APPLICANT: Searle/Monsanto
; APPLICANT: Phippard, Deborah
; APPLICANT: Vasanthakamur, Geetha
; APPLICANT: Dotson, Stanton
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,
; TITLE OF INVENTION: vectors, and cells
; FILE REFERENCE: SO-3221 PR
; CURRENT APPLICATION NUMBER: US/09/765,231A
; CURRENT FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 75
; LENGTH: 283
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-765-231A-75

Query Match 28.8%; Score 236; DB 10; Length 283;
Best Local Similarity 94.0%; Pred. No. 7.3e-54;
Matches 267; Conservative 0; Mismatches 15; Indels 2; Gaps 2;


```
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 45
; LENGTH: 2626
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 1134866.36
US-10-002-600-45

Query Match      4.6%; Score 37.4; DB 12; Length 2626;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 514 CCAGGGCTGCTGGGACGATTGCAGGTGTATGCGAGCGTTGGCTACCCACTGCCCA 573
Db 1695 CCGTGGCTTCGGGACTGGATGACAGTATCTCGCCCTTCTCAAGGTCCCTGCACAGCTT 1754
QY 574 GCCTCTGCCAGGACTGAGCCAGCGCTGGGCGCCCTGGCCCTGCCACAGTGACTTCTCTCCA 633
Db 1755 GCTTGTGATGTCCTGCTGCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1814
QY 634 GAAGATGATGACTTCTGGCTGTGAAGAGCTGAGACCTGGCTATGCGGTTCAGGCCAA 693
Db 1815 TGGGCTGACGAGGCGCGCGGGTGGAGAGCTGCAGAACCGCATCGCCAGCTGCTCTGAA 1874
QY 694 GGA 696
Db 1875 GGA 1877

RESULT 14
US-10-002-600-44
; Sequence 44, Application US/10002600
; Patent No. US2002013707A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 2670
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 1134866.26
US-10-002-600-44

Query Match      4.6%; Score 37.4; DB 12; Length 2670;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 514 CCAGGGCTGCTGGGACGATTGCAGGTGTATGCGAGCGTTGGCTACCCACTGCCCA 573
Db 1735 CCGTGGCTTCGGGACTGGATGACAGTATCTCGCCCTTCTCAAGGTCCCTGCACAGCTT 1794
QY 574 GCCTCTGCCAGGACTGAGCCAGCGCTGGGCGCCCTGGCCCTGCCACAGTGACTTCTCTCCA 633
Db 1795 GCTTGTGATGTCCTGCTGCTGCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 1854
QY 634 GAAGATGATGACTTCTGGCTGTGAAGAGCTGAGACCTGGCTATGCGGTTCAGGCCAA 693
Db 1875 GGA 1877

Db 1855 TGGGCTGCAGGAGCGCGGGTGGAGAGCTGCAGAACCGCATTCGCCAGCTGCTCTGAA 1914
QY 694 GGA 696
Db 1915 GGA 1917

RESULT 15
US-09-728-952-60/c
; Sequence 60, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 60
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (344)..(1015)
US-09-728-952-60

Query Match      4.5%; Score 36.6; DB 10; Length 1232;
Best Local Similarity 54.0%; Pred. No. 1.5;
Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 392 AGGCTGGGCTGACCCAGAACTATGAGCGCTACAGTCACTCTCTGTGTACTTGGCTGGC 451
Db 985 AGGCTCTGTCATCTCCGTCATCCCATGATTGGAGTGTCCAGTACTTGGCCCCCGCAGGCTGGC 926
QY 452 CTCACCGTCAAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGGCCCACTTCTGTACCAAGC 511
Db 925 CTGAGCCTTCAGGAAGCCCTCGCTGAATCCATCTCAGCCTGGCTGCTCTCTGACTCCCA 866
QY 512 CTCAGAGGCTGCTGGGCA 530
Db 865 GTAGAGCACTGCTGCCCA 847

Search completed: February 2, 2003, 04:00:10
Job time : 119.189 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2003, 08:38:43 ; Search time 26.3918 Seconds
(without alignments)
9516.918 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 819

Sequence: 1 tattattaaagcttcgccg.....agcccaagtcagctgtgctt 819

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819	100.0	819	1	US-08-792-019B-4
2	819	100.0	819	3	US-08-988-819-4
3	819	100.0	819	3	US-09-016-534-4
4	669.4	81.7	797	1	US-08-792-019B-1
5	669.4	81.7	797	3	US-08-988-819-1
6	669.4	81.7	797	3	US-09-016-534-1
7	627	76.6	1710	3	US-09-106-182-1
8	429.6	52.5	5087	1	US-08-792-019B-3
9	429.6	52.5	5087	3	US-08-988-819-3
10	429.6	52.5	5087	3	US-09-016-534-3
11	52	6.3	396	3	US-09-106-182-7
12	43.2	5.3	6803	3	US-08-665-259-19
13	43.2	5.3	6803	3	US-08-782-500-19
14	42.4	5.2	1642	2	US-08-665-037-1
15	42.4	5.2	1642	2	US-08-666-067-1
16	42.4	5.2	1642	2	US-08-732-870-1
17	38.4	4.7	68750	3	US-09-335-409-1
18	38.4	4.7	68750	4	US-09-568-102-1
19	38.4	4.7	68750	4	US-09-567-969-1
20	38.4	4.7	68750	4	US-09-568-480-1
21	38.4	4.7	68750	4	US-09-568-486-1
22	38.4	4.7	68750	4	US-09-568-472-1
23	38.4	4.7	68750	4	US-09-567-899-1
24	34.8	4.2	4403765	4	US-09-103-840A-2
25	34.8	4.2	4411529	4	US-09-103-840A-1
26	34.4	4.2	1225	1	US-08-197-496A-1
27	34.4	4.2	1225	1	US-08-275-370-1

Sequence 18, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 115, App
Sequence 1, Appli
Sequence 1, Appli
Sequence 18, Appli
Sequence 56, Appli
Sequence 57, Appli
Sequence 58, Appli
Sequence 56, Appli
Sequence 57, Appli
Sequence 58, Appli

28 33.6 4.1 1308 4 US-08-937-067-18
29 33.6 4.1 8147 4 US-09-514-247A-9
30 33.6 4.1 13875 2 US-08-734-344-1
31 33.6 4.1 36519 3 US-08-923-137-2
32 33.4 4.1 1117 4 US-09-347-819-5
33 33.4 4.1 1600 3 US-08-602-791-1
34 33.2 4.1 3386 4 US-09-041-886-1
35 33.2 4.1 1226 1 US-08-905-223-115
36 33 4.0 297 4 US-08-367-968-1
37 33 4.0 1226 1 US-08-665-484-1
38 32.8 4.0 261 4 US-09-450-072-18
39 32.8 4.0 261 4 US-09-351-348-18
40 32.8 4.0 1749 4 US-09-450-072-57
41 32.8 4.0 1749 4 US-09-450-072-58
42 32.8 4.0 1749 4 US-09-450-072-58
43 32.8 4.0 1749 4 US-09-351-348-56
44 32.8 4.0 1749 4 US-09-351-348-57
45 32.8 4.0 1749 4 US-09-351-348-58

ALIGNMENTS

RESULT 1
US-08-792-019B-4
; Sequence 4, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUTROTROPIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: 176..769
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..175
; US-08-792-019B-4

Query Match 100.0%; Score 819; DB 1; Length 819;
Best Local Similarity 100.0%; Pred. No. 8.3e-206; Indels 0; Gaps 0;
Matches 819; Conservative 0; Mismatches 0;

```
QY 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
QY 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 GGGGATGTTAGCTTGCATGCGCGGCTGTGCGACCTCCCTGCGAGTCCGCGAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTGCATGCGCGGCTGTGCGACCTCCCTGCGAGTCCGCGAGCTCTTAA 180
QY 181 TCGCAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 TCGCAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 CTTGGAGCATCACTCCGCGAGCTTAGTGGGAGCTTACCTGAACTACCTGGGGCCCCCTT 300
Db 241 CTTGGAGCATCACTCCGCGAGCTTAGTGGGAGCTTACCTGAACTACCTGGGGCCCCCTT 300
QY 301 CAACGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCACGGT 360
Db 301 CAACGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCACGGT 360
QY 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
QY 421 GTACAGTCACTCTCTGTGTACTTGGTGGGCTTCAACCGCTCAGGCTGCCACAGCTGAAT 480
Db 421 GTACAGTCACTCTCTGTGTACTTGGTGGGCTTCAACCGCTCAGGCTGCCACAGCTGAAT 480
QY 481 CGGAGCTAGCTGGCCCACTTCTGTATCCAGCCTCCAGGGCTGTCTGGGCGAGCATTTGAGG 540
Db 481 CGGAGCTAGCTGGCCCACTTCTGTATCCAGCCTCCAGGGCTGTCTGGGCGAGCATTTGAGG 540
QY 541 TGTCTATGGGAGCTTGGGAGTACCACTGCGCCAGCCTTCCGAGGAGTCCAGGCGCTG 600
Db 541 TGTCTATGGGAGCTTGGGAGTACCACTGCGCCAGCCTTCCGAGGAGTCCAGGCGCTG 600
QY 601 GGGCCCTGGCCCTGCCACAGTACTTCTTCCAGAGATGAGTACTTCTGGCTGCTGAA 660
Db 601 GGGCCCTGGCCCTGCCACAGTACTTCTTCCAGAGATGAGTACTTCTGGCTGCTGAA 660
QY 661 GGAGCTGACAGCTGGCTATGGCTTTCAGCCAGGAGCTTCAACCGGCTTAAAGAGAGAT 720
Db 661 GGAGCTGACAGCTGGCTATGGCTTTCAGCCAGGAGCTTCAACCGGCTTAAAGAGAGAT 720
QY 721 CGAGCTCCAGAGCTTCACTGACCTTCCAGTGGAGGACATGGTTTCTGACCTCTGAC 780
Db 721 CGAGCTCCAGAGCTTCACTGACCTTCCAGTGGAGGACATGGTTTCTGACCTCTGAC 780
QY 781 CTTTAAACCCCACTCCAGGCGCGAGTCCAGCTGCTT 819
Db 781 CTTTAAACCCCACTCCAGGCGCGAGTCCAGCTGCTT 819
```

RESULT 2

```
US-08-988-819-4
; Sequence 4, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; NAME/KEY: mat_peptide
; LOCATION: 176..769
; NAME/KEY: sig_peptide
; LOCATION: 95..175
; US-08-988-819-4
```

```
Query Match 100.0%; Score 819; DB 3; Length 819;
Best Local Similarity 100.0%; Pred. No. 8.3e-206;
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
Db 1 TATTATTAAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCGAGCCTCTGGGAGA 60
QY 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 61 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 GGGGATGTTAGCTTGCATGCGCGGCTGTGCGACCTCCCTGCGAGTCCGCGAGCTCTTAA 180
Db 121 GGGGATGTTAGCTTGCATGCGCGGCTGTGCGACCTCCCTGCGAGTCCGCGAGCTCTTAA 180
QY 181 TCGCAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 TCGCAGGAGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 CTTGGAGCATCACTCCGCGAGCTTAGTGGGAGCTTACCTGAACTACCTGGGGCCCCCTT 300
Db 241 CTTGGAGCATCACTCCGCGAGCTTAGTGGGAGCTTACCTGAACTACCTGGGGCCCCCTT 300
QY 301 CAACGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCACGGT 360
Db 301 CAACGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCACGGT 360
QY 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 361 CAACCTTGAAGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
QY 421 GTACAGTCACTCTCTGTGTACTTGGTGGGCTTCAACCGCTCAGGCTGCCACAGCTGAAT 480
Db 421 GTACAGTCACTCTCTGTGTACTTGGTGGGCTTCAACCGCTCAGGCTGCCACAGCTGAAT 480
QY 481 CCGAGCTAGCTGGCCCACTTCTGTATCCAGCCTCCAGGGCTGTCTGGGCGAGCATTTGAGG 540
Db 481 CCGAGCTAGCTGGCCCACTTCTGTATCCAGCCTCCAGGGCTGTCTGGGCGAGCATTTGAGG 540
```

QY 541 TGTCTGCGGACGCTTGGCTTACCCTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600
DB 541 TGTCTGCGGACGCTTGGCTTACCCTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600
QY 601 GGGCCCTGGCCCTGGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGAA 660
DB 601 GGGCCCTGGCCCTGGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGAA 660
QY 661 GGAGCTGCAGACCTGGCTATGGGCTTCAGCCCAAGGACTTCAACCGGCTTAAGAAGAGAT 720
DB 661 GGAGCTGCAGACCTGGCTATGGGCTTCAGCCCAAGGACTTCAACCGGCTTAAGAAGAGAT 720
QY 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGCACATGTTTCTGACCTCTGAC 780
DB 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGCACATGTTTCTGACCTCTGAC 780
QY 781 CCTTAACCCCCACACCTCCAGGCCCACTGAGCTGTGCTT 819
DB 781 CCTTAACCCCCACACCTCCAGGCCCACTGAGCTGTGCTT 819

RESULT 3

US-09-016-534-4
; Sequence 4, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 819 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..769
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 176..769
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..175
; US-09-016-534-4

Query Match 100.0%; Score 819; DB 3; Length 819;
Best Local Similarity 100.0%; Pred. No. 8.3e-206;
Matches 819; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TATTATTAAAGCTTCCGCGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGGAGA 60
DB 1 TATTATTAAAGCTTCCGCGAGCGCGGCTCGCCCTCCCACTCGGCAGGCTCTGGGAGA 60
QY 61 GGAGCCGCGCCCGCGCGCCGCCAGCCCACTGAGCTCCGAGCAGGGGACTCGTG 120
DB 61 GGAGCCGCGCCCGCGCGCCGCCAGCCCACTGAGCTCCGAGCAGGGGACTCGTG 120
QY 121 GGGGATGTTAGCTTGGCTATGCAAGTGTGTGGCACTCCCTGCACTGCAAGCTCTTAA 180
DB 121 GGGGATGTTAGCTTGGCTATGCAAGTGTGTGGCACTCCCTGCACTGCAAGCTCTTAA 180
QY 181 TCGCACAGGAGATCCAGGCCCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCGCTA 240
DB 181 TCGCACAGGAGATCCAGGCCCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCGCTA 240
QY 241 CCTGGAGATCAACTCCGAGCTTAGCTGGGACCTACTGAACTTACCTGGGGCCCCCTTT 300
DB 241 CCTGGAGATCAACTCCGAGCTTAGCTGGGACCTACTGAACTTACCTGGGGCCCCCTTT 300
QY 301 CAACGAGCTGACTTCAATCCTCTGCACTGGGGGCAAGAACTCTGCCAGGGCCACGGT 360
DB 301 CAACGAGCTGACTTCAATCCTCTGCACTGGGGGCAAGAACTCTGCCAGGGCCACGGT 360
QY 361 CAACCTTGGAAAGTGGCGAAGCTCAATGACAGGCTGGGCTGACCCAGAACATATGAGGC 420
DB 361 CAACCTTGGAAAGTGGCGAAGCTCAATGACAGGCTGGGCTGACCCAGAACATATGAGGC 420
QY 421 GTACAGTCACTCTCTGTGTTACTTGGTGGCTCAACCGCTCAGGCTGCACAGCTGAAT 480
DB 421 GTACAGTCACTCTCTGTGTTACTTGGTGGCTCAACCGCTCAGGCTGCACAGCTGAAT 480
QY 481 CCGAGCTAGCTGGCCCACTTCTGTACAGCTCCAGGGCTGTCTGGGCAAGATTCGAGG 540
DB 481 CCGAGCTAGCTGGCCCACTTCTGTACAGCTCCAGGGCTGTCTGGGCAAGATTCGAGG 540
QY 541 TGTCTGCGGACGCTTGGCTACCCTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600
DB 541 TGTCTGCGGACGCTTGGCTACCCTGCCCCAGCCCTCTGCCAGGACTGAGCCAGCCTG 600
QY 601 GGGCCCTGGCCCTGGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGAA 660
DB 601 GGGCCCTGGCCCTGGCCACAGTGAATCTTCCAGAGATGGATGACTTCTGGCTGTGAA 660
QY 661 GGAGCTGCAGACCTGGCTATGGGCTTCAGCCCAAGGACTTCAACCGGCTTAAGAAGAGAT 720
DB 661 GGAGCTGCAGACCTGGCTATGGGCTTCAGCCCAAGGACTTCAACCGGCTTAAGAAGAGAT 720
QY 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGCACATGTTTCTGACCTCTGAC 780
DB 721 GCAGCTCCAGCAGCTTCAGTCACTCCCTGCACTTGGAGGCACATGTTTCTGACCTCTGAC 780
QY 781 CCTTAACCCCCACACCTCCAGGCCCACTGAGCTGTGCTT 819
DB 781 CCTTAACCCCCACACCTCCAGGCCCACTGAGCTGTGCTT 819

RESULT 4

US-08-792-019B-1
; Sequence 1, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS

```
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER TYPE: Floppy disk
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-08-792-019B-1

Query Match      81.7%; Score 669.4; DB 1; Length 797;
Best Local Similarity 92.0%; Pred. No. 1.4e-166;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 5 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCTGGGAGAGGAG 64
Db 1 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCTGGGAGAGGAG 60
QY 65 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 124
Db 61 CGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
QY 125 ATGTTAGCTTGCTATGACGCGTGCTGCGACCTCCCTGCGAGTCCAGCTCTTAATCGC 184
Db 120 ATGTTAGCTTGCTATGACGCGTGCTGCGACCTCCCTGCGAGTCCAGCTCTTAATCGC 179
QY 185 ACAGGAGATCCAGGCGCTCGCGCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTG 244
Db 180 ACAGGGGACCCAGGCGCTCGCGCGCTCCATCCAGAAACCTATGACCTCACCGCTACCTG 239
QY 245 GAGCATCACTCGGAGCTTAGCTGGGAGCTTACCTGAACCTACCTGGGGCCCCCTTTCAAC 304
Db 240 GAGCACCAACTCGGAGCTTGCGTGGGAGCTTCTGAACCTACCTGGGCCCCCTTTCAAC 299
QY 305 GAGCCTGACTTCAATCCTCTCGACTGGGCGGAGAACTCTGCCAGGGCCACGGTCAAC 364
Db 300 GAGCCAGACTTCAACCTCTCCGCGCTGGGGCGGAGAGCTCTGCCAGGGCCACTGTGAC 359
QY 365 TTGGAAGTGTGGGAGAGCTCAATGACAGGCTCGCGCTGACCCAGAACTATGAGGCGTAC 424
Db 360 TTGGAAGTGTGGGAGAGCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGGCGTAC 419
QY 425 AGTCACCTCTCTGTGTACTTGTGGTGGCGCTCAACCGTCAGCTGCCAGCTGAACTCCGA 484
Db 420 AGCCACCTTCTGTGTACTTGTGGTGGCGCTCAACCGTCAGCTGCCAGCTGAACTCCGA 479
QY 485 CGTAGCTGGCGGCACTTCTGTATCCAGCGCTTCCAGGGGCTGCTGGGCGAGCAATTGAGGTGC 544
Db 480 CGACGCTGGCGGCACTTCTGTATCCAGCGCTTCCAGGGGCTGCTGGGCGAGCAATTGAGGTGC 539
```

```
QY 545 ATGGCGAGCGCTTGCTACCCACTGCCAGGCTCTCCAGGAGCTGAGCCAGCTGGGCC 604
Db 540 ATGGCAGCTCTGGGCTACCCACTGCCAGCGCTGCCCTGGGACTGAACCCACTTGGACT 599
QY 605 CTTGGCCCTGCCACAGTGAATCTCTCCAGAAATGAGTGAATCTTGGCTGTGAAGGAG 664
Db 600 CTTGGCCCTGCCACAGTGAATCTCTCCAGAAATGAGTGAATCTTGGCTGTGAAGGAG 659
QY 665 CTGAGACCTGGCTATGCGTTGAGCCAGGAGCTTCAACCGCTTAAGAGAGATGCGAG 724
Db 660 CTGAGACCTGGCTATGCGTTGAGCCAGGAGCTTCAACCGCTTAAGAGAGATGCGAG 719
QY 725 CTTCCAGCAGCTTTCAGTCCACCTGACCTTGAGGAGCAGATGTTCTGACCTCTGACCT 783
Db 720 CTTCCAGCAGCTTTCAGTCCACCTGACCTTGAGGAGCAGATGTTCTGACCTCTGACCT 778
```

```
RESULT 5
US-08-988-819-1
; Sequence 1, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUTROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 90..764
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 171..764
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 90..170
; US-08-988-819-1
```

```
Query Match      81.7%; Score 669.4; DB 3; Length 797;
Best Local Similarity 92.0%; Pred. No. 1.4e-166;
Matches 717; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

QY 5 ATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCGCCAGAGCTCTGGGAGAGGAG 64
```



```
/ FILING DATE: 03-FEB-1997
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: COOK, ROBERT R.
/ REGISTRATION NUMBER: 31,602
/ REFERENCE/DOCKET NUMBER: A-442
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5087 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 137-138
/ OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED
/ OTHER INFORMATION: REGION OF >1KB"
US-08-792-019B-3

Query Match 52.5%; Score 429.6; DB 1; Length 5087;
Best Local Similarity 90.4%; Pred. No. 1.8e-103;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 335
DB 3363 AGCTGAACCTACCTGGGGCCCCCTTTCAACGAGCGAGACTTCAACCTCCCGCTGGGG 3422
QY 336 CAGNAACCTGCCAGGCGCCACGCTCACTTGGAGCTGGCGAGGCTCAATGACAGGC 395
DB 3423 CAGAGACTCTGCCAGGCGCCACTTTCACCTTGGAGGTGGCGAAGGCTCAATGACAAAC 3482
QY 396 TGGCGCTGACCCAGAACTATGAGCGCTACAGTCACTCTCTGTGTACTTGGCTGGCTCA 455
DB 3483 TGGCGCTGACCCAGAACTATGAGCGCTACAGGCGCTTCTGTGTACTTGGCTGGCTCA 3542
QY 456 ACCGTCAAGGCTGCACAGCTGAACTCCGAGCTAGCTGGGCCCACTTCTGTACCAAGCTCC 515
DB 3543 ACCGTCAAGGCTGCACAGCTGAACTCCGAGCTAGCTGGGCCCACTTCTGTACCAAGCTCC 3602
QY 516 AGGGCTCTGGGCGAGCTTGCAGTGTGCTGCGAGCTTGGCTAGCCCTACCCCTGCCAGC 575
DB 3603 AGGGCTCTGGGCGAGCTTGCAGTGTGCTGCGAGCTTGGCTAGCCCTACCCCTGCCAGC 3662
QY 576 CTCTGCAGGAGCTGAGCCAGCTTGGGCCCCCTTGGCCCCCTGCCCCAGTCTTCTCCAGA 635
DB 3663 CGTGTCTGGGAGCTGAACCCACTTGGACTCTTGGCCCCCTGCCCCAGTCTTCTCCAGA 3722
QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCACTGACCTGCACTGG 755
DB 3723 AGATGGAGCACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCACTGACCTGCACTGG 3782
QY 696 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCACTGACCTGCACTGG 755
DB 3783 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCACTGACCTGCACTGG 3842
QY 756 AGGCACATGTTTCTGACCTCTGACCTT 783
DB 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

RESULT 9
US-08-988-819-3
/ Sequence 3, Application US/08988819
/ Patent No. 6054294
/ GENERAL INFORMATION:
/ APPLICANT: CHANG, MING-SHI
/ TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: AMGEN INC.
/ STREET: ONE AMGEN CENTER DRIVE
/ CITY: THOUSAND OAKS
```

```
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 91320
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/988,819
/ FILING DATE: 12-DEC-1997
/ CLASSIFICATION: 536
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/792,019
/ FILING DATE: 03-FEB-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: COOK, ROBERT R.
/ REGISTRATION NUMBER: 31,602
/ REFERENCE/DOCKET NUMBER: A-442A
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5087 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 137-138
/ OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED
/ OTHER INFORMATION: REGION OF >1KB"
US-08-988-819-3

Query Match 52.5%; Score 429.6; DB 3; Length 5087;
Best Local Similarity 90.4%; Pred. No. 1.8e-103;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 276 ACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 335
DB 3363 AGCTGAACCTACCTGGGGCCCCCTTTCAACGAGCGAGACTTCAACCTCCCGCTGGGG 3422
QY 336 CAGNAACCTGCCAGGCGCCACGCTCACTTGGAGCTGGCGAGGCTCAATGACAGGC 395
DB 3423 CAGAGACTCTGCCAGGCGCCACTTTCACCTTGGAGGTGGCGAAGGCTCAATGACAAAC 3482
QY 396 TGGCGCTGACCCAGAACTATGAGCGCTACAGTCACTCTCTGTGTACTTGGCTGGCTCA 455
DB 3483 TGGCGCTGACCCAGAACTATGAGCGCTACAGGCGCTTCTGTGTACTTGGCTGGCTCA 3542
QY 456 ACCGTCAAGGCTGCACAGCTGAACTCCGAGCTAGCTGGGCCCACTTCTGTACCAAGCTCC 515
DB 3543 ACCGTCAAGGCTGCACAGCTGAACTCCGAGCTAGCTGGGCCCACTTCTGTACCAAGCTCC 3602
QY 516 AGGGCTCTGGGCGAGCTTGCAGTGTGCTGCGAGCTTGGCTAGCCCTACCCCTGCCAGC 575
DB 3603 AGGGCTCTGGGCGAGCTTGCAGTGTGCTGCGAGCTTGGCTAGCCCTACCCCTGCCAGC 3662
QY 576 CTCTGCAGGAGCTGAGCCAGCTTGGGCCCCCTTGGCCCCCTGCCCCAGTCTTCTCCAGA 635
DB 3663 CGTGTCTGGGAGCTGAACCCACTTGGACTCTTGGCCCCCTGCCCCAGTCTTCTCCAGA 3722
QY 636 AGATGGATGACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCACTGACCTGCACTGG 695
DB 3723 AGATGGAGCACTTCTGGCTGCTGAAGAGCTGCGAGCTGCGAGCTTCACTGACCTGCACTGG 3782
QY 696 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCACTGACCTGCACTGG 755
DB 3783 ACTTCAACCGGCTTAAAGAGAGTGCAGCTCCAGAGCTTCACTGACCTGCACTGG 3842
QY 756 AGGCACATGTTTCTGACCTCTGACCTT 783
DB 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870
```

RESULT 10

US-09-016-534-3
; Sequence 3, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARNALDI, ULLA
; APPLICANT: SARNALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 137..138
; OTHER INFORMATION: /product= "INTERVENING UNSEQUENCED"
; OTHER INFORMATION: REGION OF >1KB"

US-09-016-534-3

Query Match 52.5%; Score 429.6; DB 3; Length 5087;
Best Local Similarity 90.4%; Pred. No. 1.8e-103;
Matches 459; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 276 ACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCTGACCTTCAATCTCTCGACTGGGGG 335
Db 3363 AGCTGAACCTACCTGGGGCCCCCTTTCAACGAGCTGACCTTCAATCTCTCGACTGGGGG 3422
QY 336 CAGAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCTCAATGACAGGC 395
Db 3423 CAGAGACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCTCAATGACAAAC 3482
QY 396 TCGGGTGACCCAGAACTATGAGCGGTACAGTCACCTCTCTGTGTACTTTCGTGGCTCA 455
Db 3483 TCGGGTGACCCAGAACTACGAGCGCTACAGCCACCTCTCTGTGTACTTTCGTGGCTCA 3542
QY 456 ACCGTGAGCTGCCAGCTGACCTCGAGGTAGCTTGGCCCACTTCTGTACAGCTCC 515
Db 3543 ACCGTGAGCTGCCAGCTGACCTCGAGGTAGCTTGGCCCACTTCTGTACAGCTCC 3602
QY 516 AGGGCTGTCTGGGCGACATTGAGGTGTCTATGGCGACGCTTGGCTACCCACTGCCCGCAGC 575
Db 3603 AGGGCTGTCTGGGCGACATTGAGGTGTCTATGGCGACGCTTGGCTACCCACTGCCCGCAGC 3662

QY 576 CTCTGCCAGGAGCTGAGCCAGCTGGGGCCCCCTGGCCCTGCCACACAGTGACTTCTCTCAGA 635
Db 3663 CGCTGGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCACACAGTGACTTCTCTCAGA 3722
QY 636 AGATGGATGACTTCTGGCTGTCTGAAGAGCTGCAGACCTGGCTATGGCCCTTACAGCCAAGG 695
Db 3723 AGATGGAGCAGCTTCTGGCTGTCTGAAGAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGG 3782
QY 696 ACTTCAACCGGCTTAAGAGAGATGCAGCTCCAGCAGCTTCACTCACCCTGCACCTTGG 755
Db 3783 ACTTCAACCGGCTCAAGAAAGATGCAGCTCCAGCAGCTTCACTCACCCTGCACCTTGG 3842
QY 756 AGGCACATGGTTTCTGACCTCTGACCCCT 783
Db 3843 GGGCTCATGGCTTCTGACTTCTGACCTT 3870

RESULT 11

US-09-106-182-7
; Sequence 7, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yanggu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,182
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/051,053
; FILING DATE: 30-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-106-182-7

Query Match 6.3%; Score 52; DB 3; Length 396;
Best Local Similarity 85.3%; Pred. No. 8e-05;
Matches 59; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 716 AAGATGAGCCCTCCAGCAGCTTCAGTCACTTGGAGGCAATGGTTTCTGACCT 775
Db 4 ACGAGCAGCCCTCCAGCAGCTTCAGTCACTTGGAGGCTCATGGCTTCTAACTT 63
QY 776 CTGACCTT 783
Db 64 CTGACCTT 71

RESULT 12

US-08-665-259-19/c
; Sequence 19, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-665-259-19

Query Match 5.3%; Score 43.2; DB 3; Length 6803;
Best Local Similarity 49.6%; Pred. No. 0.038;
Matches 111; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 24 CGCGCTCGCCCTCCACTCCGCGCAGCCTCTGGGAGAGGAGCGCGCGCGCGCGCGCG 83
DB 6708 CCGCGCCCCACGCTCCCGCCCCACGACGCGCGCGCGCGCGCGCGCGCGCGCG 6649
QY 84 CCCCAGCCCCATGACCTCCGAGCAGGGGACTCTGTGGGGGATGTTAGCTTTCCTATGCA 143
DB 6648 CCGCGCACCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6589
QY 144 CGGTGCTGTGGACCTCCCTGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCCCTG 203
DB 6588 CGCGCTTTCCTCAGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6529
QY 204 GCCCTCCATCCAGAAACCTATGACCTCACCGCTACTCGAG 247
DB 6528 CCGCGCGAGCTGTGCGCTGTGCCACCGCGGAGAGCGCGCG 6485

RESULT 13

US-08-762-500-19/c
; Sequence 19, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-762-500-19

Query Match 5.3%; Score 43.2; DB 3; Length 6803;
Best Local Similarity 49.6%; Pred. No. 0.038;
Matches 111; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 24 CGCGCTCGCCCTCCACTCCGCGCAGCCTCTGGGAGAGGAGCGCGCGCGCGCGCG 83
DB 6708 CCGCGCCCCACGCTCCCGCCCCACGACGCGCGCGCGCGCGCGCGCGCGCGCG 6649
QY 84 CCCCAGCCCCATGACCTCCGAGCAGGGGACTCTGTGGGGGATGTTAGCTTTCCTATGCA 143
DB 6648 CCGCGCACCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6589
QY 144 CGGTGCTGTGGACCTCCCTGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCCCTG 203
DB 6588 CGCGCTTTCCTCAGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6529
QY 204 GCCCTCCATCCAGAAACCTATGACCTCACCGCTACTCGAG 247
DB 6528 CCGCGCGAGCTGTGCGCTGTGCCACCGCGGAGAGCGCGCG 6485

RESULT 14

US-08-665-037-1
; Sequence 1, Application US/08665037
; Patent No. 5895813
; GENERAL INFORMATION:

APPLICANT: Seedorf, Klaus
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF TKA-1 RELATED
TITLE OF INVENTION: DISORDERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,037
FILING DATE: June 13, 1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,167
FILING DATE: October 13, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 220/156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1642 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 97...1446
US-08-665-037-1

Query Match 5.2%; Score 42.4; DB 2; Length 1642;
Best Local Similarity 52.9%; Pred. No. 0.04; Mismatches 0; Indels 0; Gaps 0;
Matches 91; Conservative 0;

QY 512 CTCACGGGCTGCTGGGCGAGCATTCAGAGTGTCATGGCGAGCTTGCTGCTACCCACTGCC 571
DB 379 CTCGCGCGGCGGAGCTGACCTGTACCGAGGAGATGGCCCGAGGAGGCTCCACCGCC 438
QY 572 CAGCCTCTCCAGGAGTACGAGCGAGCTGGGCGCCCTGGCCCTGCCACAGTACTTCCTC 631
DB 439 CACGACCCCTGGGAGCGGAGCGAGCTGGGCGACACACCGCGGAGGAGCTCCGAGCT 498
QY 632 CAGAAGATGATGACTTCTGGCTGTGAAGGAGCTGAGAGCTGAGCTATGCG 683
DB 499 GCGAAGAAGATGATGAGTGGGCGCCCTGAGGAGCTGGCGCCCTCGGCTCTGCC 550

RESULT 15
US-08-666-067-1
Sequence 1, Application US/08666067
Patent No. 5922842
GENERAL INFORMATION:
APPLICANT: Seedorf, Klaus
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF TKA-1 RELATED
TITLE OF INVENTION: DISORDERS

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/666,067
FILING DATE: June 13, 1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,421
FILING DATE: October 13, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 220/157
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1642 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 97...1446
US-08-666-067-1

Query Match 5.2%; Score 42.4; DB 2; Length 1642;
Best Local Similarity 52.9%; Pred. No. 0.04; Mismatches 0; Indels 0; Gaps 0;
Matches 91; Conservative 0;

QY 512 CTCACGGGCTGCTGGGCGAGCATTCAGAGTGTCATGGCGAGCTTGCTGCTACCCACTGCC 571
DB 379 CTCGCGCGGCGGAGCTGACCTGTACCGAGGAGATGGCCCGAGGAGGCTCCACCGCC 438
QY 572 CAGCCTCTGCCAGGAGTACGAGCGAGCTGGGCGCCCTGGCCCTGCCACAGTACTTCCTC 631
DB 439 CACGACCCCTGGGAGCGGAGCGAGCTGGGCGACACACCGCGGAGGAGCTCCGAGCT 498
QY 632 CAGAAGATGATGACTTCTGGCTGTGAAGGAGCTGAGAGCTGAGCTATGCG 683
DB 499 GCGAAGAAGATGATGAGTGGGCGCCCTGAGGAGCTGGCGCCCTCGGCTCTGCC 550

Search completed: February 1, 2003, 08:50:20
Job time : 46.3918 secs

```

RESULT 2
49153
cardiotrophin-1 - mouse
Species: Mus musculus (house mouse)
Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
Accession: I49153
Fennica, D.; King, K.L.; Shaw, K.J.; Luis, E.; Rullamas, J.; Luch, S.; Dar-
roc. Natl. Acad. Sci. U.S.A. 92, 1142-1146, 1995
Title: Expression cloning of cardiotrophin 1, a cytokine that induces card-
Reference number: I49153; MUID:95166785; PMID:7862649
Accession: I49153
Status: nucleic acid sequence not shown; translated from GE/EMBL/DBBJ
Molecule type: mRNA
Residues: 1-203 <RES>
Cross-References: EMBL:U18366; NID:G710331; PID:G710332
Genetics:
Gene: ctfl

```


A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54231.1; PID:gl7985203; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10989
A:Map position: II

Query Match 7.6%; Score 92.5; DB 2; Length 200;
Best Local Similarity 23.9%; Pred. No. 0.39;
Matches 55; Conservative 21; Mismatches 89; Indels 65; Gaps 11;

QY 10 GMLACLTVLWHLPAVPAALNRTGDPGPGPSIQKTYDLYTRYLEHQVRSAGTYNLYLGPFF 69
DB 2 GLLAALV---LPSLPAEAKTQQAAMPNATSPHQADVYL---LRGFADISTGI--- 51
QY 70 NEPDFNPRLGAETLPRTATNLEV-----WRLNDRRLRLTQNYEAYSHLLCYLRLGNROA 124
DB 52 -----DEIGAE-LQAAGVNAHVQGHAAWRLVLRVADQOKNGHLPVVLIGHSLGANA 103
QY 125 A---TAELRSLAHFCTSLQGLGSLAGVMAATLGYPLPQPLPGT-----EPAW--- 169
DB 104 AIYIAEELERR-----GLAVDYMATFAATGDPDLEGNVVRVNVFYFKQHGWL 151
QY 170 --APGPAHSDFLQKMDDFWLKELQTLWLRSAKDFNRLK-KMQQPPAAV 216
DB 152 PLVGPFRPHGLENRD-----FSNAKGVGHFNIEKQRLQAEV 189

RESULT 5
G86182
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86182
R:Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor-
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Liu, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86182
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <STO>
A:Cross-references: GB:AE005172; NID:g7211973; PIDN:AAF40444.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 7.5%; Score 92; DB 2; Length 332;
Best Local Similarity 25.9%; Pred. No. 0.79;
Matches 57; Conservative 24; Mismatches 73; Indels 66; Gaps 12;

QY 23 PAVPALNRTGDPGPGPSIQKTYDLYTRYLEHQVRSAGTYNLYL--- 64
DB 113 PSVTAGNLSGYP--PRFSF--TYDPGYEQRMESLLOQFIRERNPQIRPLRLGLGSPVG 169
QY 65 LGPPFPNEPFPNPLGAETLPRTATNLEVWRLNDRRLRLTQNYEAYSHLLCYLRLGNROA 124
DB 170 LGPIRASQPLQPRVA---PPPTSILDTSRNRKRSK-----DGALAVVRG--RKV 215
QY 125 ATAEELRSL-----AHFCTSLQGLGSLAGVMAATLGYPLPQPLP---GTEPAAW 170
DB 216 RITEGSSSSYLSGRSLKNGAHV-----GIQPKRSKIMK-----PUPKPLVDLTITSTVP 266
QY 171 PGPAHSDFLQKMDDFWLKELQTLWLRSAKDFNRLKMQ 210

C;Keywords: cytokine; glycoprotein
F;164/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.0%; Score 110.5; DB 2; Length 203;
Best Local Similarity 27.4%; Pred. No. 0.0084;
Matches 49; Conservative 23; Mismatches 86; Indels 21; Gaps 6;

QY 40 IQKTYDLYTRYLEHQVRSAGTYNLYLGPFFNEPFPNPLP---GAETLPRTATNLEVWRS 96
DB 27 IRQTNLARLLTKYARQLLEEVVQQGEPFGLPGFPPRLPLAGLSGPAPSHAGLPV--- 83
QY 97 LNDRLRLTQNYEAYSHLLCYLRLGNROA-----TAELRSLAHFCTSLQGLGSLAGV 151
DB 84 -SERLR--QDAALSVLPALLDAVRRQRQELNPRAPRLRLSLEDAARQVRAAAVETVL 140
QY 152 ATLGYPLPQPLPCTEPA-----WAPGPAHSDFLQKMDDFWLKELQTLWLRSAKDFNRL 205
DB 141 AALGAAARG--POPEVTVATLFTANSTAGIFSAKVLGFHVCGLYGEWVSRTEGDLQL 197

RESULT 3
T34916
transferase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 02-Sep-2000
C:Accession: T34916
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z21558
A:Accession: T34916
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-640 <OLI>
A:Cross-references: EMBL:AL021409; PIDN:CAAL16181.1; GSPDB:GN000070; SCOEDB:SC3F7.10
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC3F7.10
C:Superfamily: glycine C-acetyltransferase homology
F;287-624/Domain: glycine C-acetyltransferase homology <GCA>

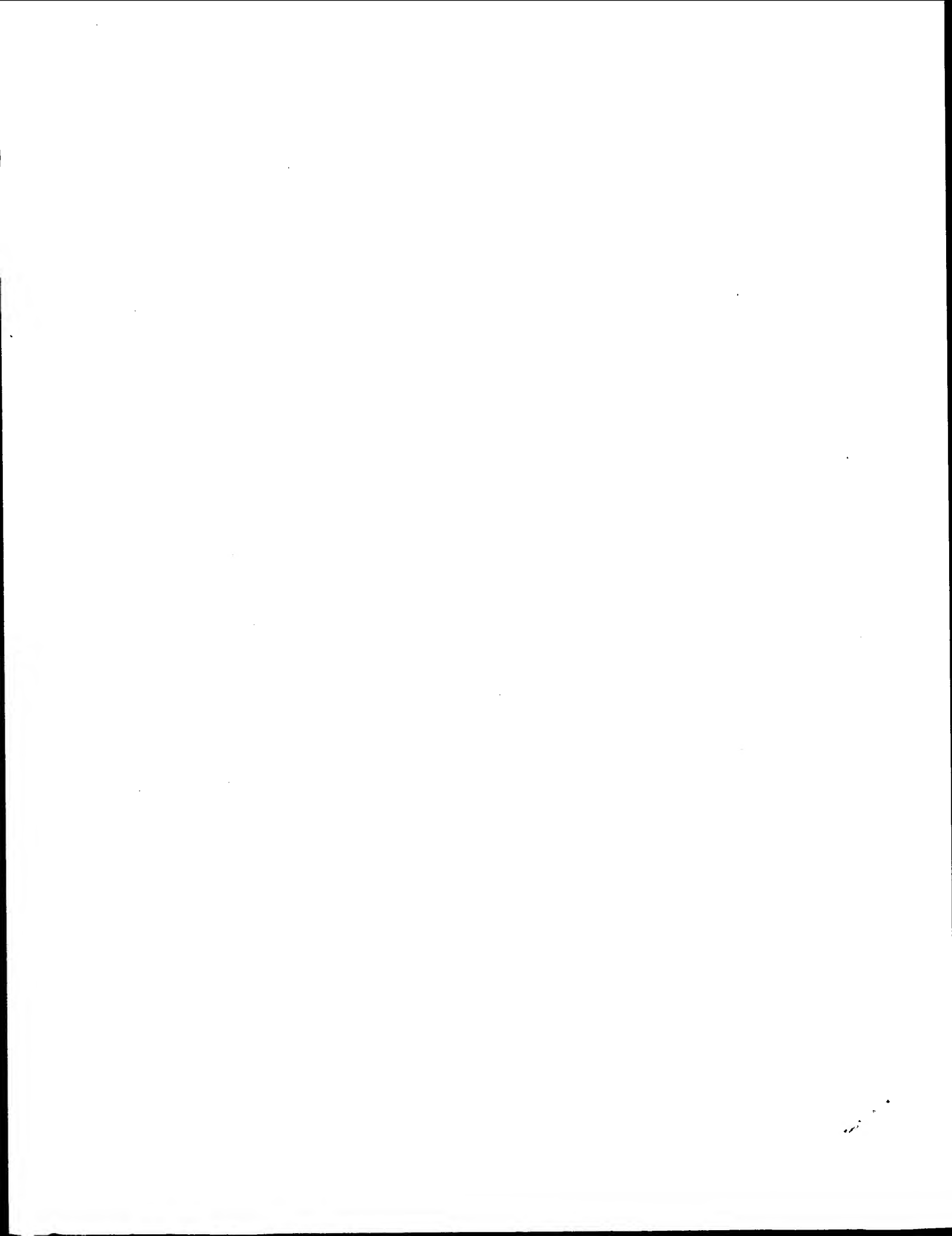
Query Match 7.6%; Score 93.5; DB 2; Length 640;
Best Local Similarity 28.3%; Pred. No. 1.3;
Matches 52; Conservative 12; Mismatches 63; Indels 57; Gaps 9;

QY 22 LPAPALNRTGDPGPGPSIQKTYDLYTRYLEHQV---RSLAGTYNLYLGPFFNEPFPNPR 78
DB 100 LPAPVA-EGTCDPVTAEVAAAMRQTQYRQHLGLDADLEG-----E 140
QY 79 LGAEETLPRTATNLEVWRLNDRRLRLTQNYEAYSHLLCYLRLGNROAATAEELRSLAHFCT 138
DB 141 LGVDSVVLTSVVAEA-----TERLGLT-----GNAAPDAAGATTILALA---D 179
QY 139 SLOGLGSLAGVMAATLGYPLPQPLFGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWLR 198
DB 180 ALRGL-----VAAAPGTVPEAPATGAA-APAPGRSGNAP-----APGADGWDHRS 225
QY 199 AKDF 202
DB 226 MKDF 229

RESULT 4
AD3633
hypothetical protein BMEI10989 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3633
R:DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3633

[illegible]

Search completed: January 27, 2003, 15:42:36
Job time : 16.5 secs



CC similarity).

CC -!- TISSUE SPECIFICITY: Expressed in brain, testis, kidney, stomach,

CC small intestine, liver, and lung. Not detected in heart, skeletal

CC muscle, and spleen.

CC -!- MISCELLANEOUS: Mice homozygous for a null mutation of the CST gene

CC display hindlimb weakness from week 6 of age and subsequently show

CC a pronounced tremor and progressive ataxia. Myelin vacuolation is

CC observed in the cerebellar white matter, diencephalon, brainstem

CC and spinal anterior column. Male mice were infertile due to a

CC blocked spermatogenesis.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; AB032940; BAA93009.1; -

CC EMBL; AB032939; BAA93008.1; -

CC EMBL; AK007645; BAB25160.1; -

CC EMBL; BC026806; AAH26806.1; -

CC DR MGD; MGI:1858277; Gcst.

CC Transferase; Transmembrane; Glycoprotein.

CC DOMAIN 1 12

CC TRANSMEM 13 35

CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

CC (POTENTIAL).

CC DOMAIN 36 423

CC LUMENAL, CATALYTIC (POTENTIAL).

CC CARBOHYD 66 66

CC N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 312 312

CC N-LINKED (GLCNAC. .) (POTENTIAL).

CC CONFLICT 18 18

CC L -> P (IN REF. 2).

CC CONFLICT 263 263

CC R -> Q (IN REF. 2).

CC CONFLICT 271 271

CC E -> K (IN REF. 2).

CC CONFLICT 358 358

CC Q -> R (IN REF. 1).

CC CONFLICT 392 392

CC R -> G (IN REF. 2).

CC CONFLICT 398 398

CC I -> T (IN REF. 1).

CC SEQUENCE 423 AA; 48968 MW; FD54A1A71F4ABE46 CRC64;

CC

CC Query Match 7.6%; Score 93; DB 1; Length 423;

CC Best Local Similarity 23.1%; Pred. No. 0.32;

CC Matches 55; Conservative 22; Mismatches 65; Indels 96; Gaps 11;

QY 44 YDLTRYLHQLRSAGTYLNY---LGPP-----FNEP----- 72

Db 204 YDPSSNAHYLRNLLFFDLGYDSSLDPASPRVQEHILEVERFHLVLLQYFDESILVLR 263

QY 73 -----DFNPRLGATLPRATVNLVWVSLNDRRLTQNYEAYSIL 113

Db 264 ELLCWDLEDVLYFKLNARDSPVPLSGELYRRAT-----AMNLD--VRLYRHFNASFWR 317

QY 114 LCYLRLGNRQA-ATAELRRS---LAHFCTSLQGLLGSIA-----GWMATLGYP 157

Db 318 KVEAFGRMARVAELQANEHMRHCIDGQAVGAERAIQDSAMPQWPLGIKSLGVN 377

QY 158 LPQPLPGTEPAWPGPAHSDFLOKMDDFWLLKELQ-----TWLWRSKDFNR 204

Db 378 LKKSII-----GPOHEQLCRM-----LTPEIQYLSDLGANLWVTKLWFLRDLR 422

RESULT 4

CTF1 HUMAN

ID_CTF1 HUMAN STANDARD; PRT; 201 AA.

AC Q16619;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cardiotrophin-1 (CT-1).

GN CTF1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.

RP TISSUE-Heart;

RX MEDLINE=96429882; PubMed=8833032;

RA Pennica D., Swanson T.A., Shaw K.J., Kuang W.-J., Gray C.L.,

RA Beatty B.G., Wood W.I.;

RT "Human cardiotrophin-1: protein and gene structure, biological and

RL binding activities, and chromosomal localization.";

RL Cytokine 8:183-189(1996)

CC -!- FUNCTION: INDUCES CARDIAC MYOCYTE HYPERTROPHY IN VITRO. BINDS TO

CC AND ACTIVATES THE LEUKEMIA INHIBITORY FACTOR RECEPTOR (LIF

CC RECEPTOR)/GP 130 RECEPTOR COMPLEX.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,

CC PROSTATE AND OVARY. LOWER LEVELS IN LUNG, KIDNEY, PANCREAS,

CC THYMUS, TESTIS AND SMALL INTESTINE. LITTLE OR NO EXPRESSION IN

CC BRAIN, PLACENTA, LIVER, SPLEEN, COLON OR PERIPHERAL BLOOD

CC LEUKOCYTES.

CC -!- SIMILARITY: BELONGS TO THE IL-6 SUPERFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; U43033; AAD12173.1; -

CC EMBL; U43031; AAD12173.1; JOINED.

CC EMBL; U43032; AAD12173.1; JOINED.

CC EMBL; U43030; AAA85229.1; -

CC DR Genew; HGNC:2499; CTF1.

CC MIM; 600435; -

CC Cytokine.

CC SEQUENCE 201 AA; 21227 MW; 0235A7B5745F675F CRC64;

CC

CC Query Match 7.5%; Score 91.5; DB 1; Length 201;

CC Best Local Similarity 25.7%; Pred. No. 0.17;

CC Matches 45; Conservative 23; Mismatches 15; Gaps 5;

QY 40 IQKTVDTRYLHQLRSAGTYLNYLGGPFNPPFNPRLGATLPRATVNLVWVSLND 99

Db 27 IRQTHSLAHLTKYAEQLLQYVQLQGDGPFGLPSPRLPVAGL-SAPAPSHAGLPVHE 85

QY 100 RLRLTQNYEAYSHLQYLRGLNRQA-----TAEIIRSLAHFCTSLQGLLGSIAWATL 154

Db 86 RLRL--DAALAAALPFLDVAVCRRQAEINPRAPRLRLRLLEDAARQAALGAAVEALLAAL 143

QY 155 GY-----PLPQPLPGTEPAWPGPAHSDFLOKMDDFWLLKELQTLWRSKDFNRL 205

Db 144 GAANRGPRAEPPAATASA---ASATGVFPAKVLGVLCVGLVRLSRTGDLGQL 195

RESULT 5

CNTF CHICK

ID_CNTF CHICK STANDARD; PRT; 195 AA.

AC Q02011;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Ciliary neurotrophic factor (CNTF) (Growth promoting activity) (GPA).

GN CNF OR GPA.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN SEQUENCE FROM N.A.

RP MEDLINE=92304573; PubMed=1610564;

RX Leung D.W., Parent A.S., Cachianes G., Lee A.L., Nikolics K.,

RA Esch F., Coulombe J.N., Blacher R.W., Eckenstein F.P., Nishi R.;

RT "Cloning, expression during development, and evidence for release of
 RT a trophic factor for ciliary ganglion neurons.";
 RL Neuron 8:1045-1053(1992).
 CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
 CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
 CC AXOTOMY.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M80827; AAA48784.1; -;
 DR PIR; JH0680; JH0680.
 DR HSPF; P26441; ICNT.
 DR InterPro; IPR000151; Ctl_neuro_factor.
 DR Pfam; PF01110; CNTF; 1.
 DR ProDom; PD011041; Ctl_neuro_factor; 1.
 DR Growth factor; Neurone.
 KW SEQUENCE 195 AA; 21330 MW; FEA076949DB34AC5 CRC64;
 SQ
 Query Match 7.4%; Score 90; DB 1; Length 195;
 Best Local Similarity 27.7%; Pred. No. 0.23;
 Matches 52; Conservative 21; Mismatches 83; Indels 32; Gaps 10;
 QY 46 LTRYLEHQRLSLAGTYLVNLPFPNPPPLGAEITLPRATVNLVWRSNDRLRLTQ 105
 DB 23 LARKMRSDVTDLDIYVERQG-----LDASISVAADVGVPTAAV--ERMAEQTQRLLD 75
 QY 106 N---YEAYSHLLCYLRLGNLROA---ATAELRRSLA-----HFCTSLQGLLGSIGAWMA 152
 DB 76 NLAAVPAFTLLAQMLEEQRELLGDDTDAELGPALAAWLLQSAFVYHLELL-----ELE 130
 QY 153 TLGYPLPQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWLRSAKDFNRLLKKMQPP 212
 DB 131 SRGAPAE---GSEPP-AP-PRLSLFEQKLRGLRLVRLAQWAVRSVRDLRLQLSK--HGP 183
 QY 213 AASVTLHL 220
 DB 184 GSGAALGL 191
 RESULT 6
 M2B2 HUMAN
 ID M2B2 HUMAN STANDARD; PRT; 1009 AA.
 AC Q9Y2E5;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Epididymis-specific alpha-mannosidase precursor (EC 3.2.1.24)
 DE (Mannosidase alpha class 2B member 2).
 GN MAN2B2 OR KIA00935.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stone N.E., Schmutz J.J., Cox D.R., Myers R.M.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 132-1009 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9924603; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.

The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 DNA Res. 6:63-70(1999).
 [3]
 RECONSTRUCTION FROM GENOMIC SEQUENCE.
 RA Bairoch A.;
 RL Unpublished observations (NOV-2001).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
 CC mannose residues in alpha-D-mannosides.
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AC004480; -; NOT ANNOTATED CDS.
 DR EMBL; AB023152; BAA76779.1; ALT SEQ.
 KW Hydrolase; Glycosidase; Signal; Glycoprotein.
 FT SIGNAL 1 23
 FT CHAIN 24 1009
 FT EPIDIDYMIS-SPECIFIC ALPHA-MANNOSIDASE.
 FT CARBOHYD 226 226
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 249 249
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 294 294
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 336 336
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 516 516
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 608 608
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 670 670
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 675 675
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 748 748
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 808 808
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 812 812
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 890 890
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1009 AA; 113987 MW; F3DB81DD061352E6 CRC64;
 Query Match 7.2%; Score 88; DB 1; Length 1009;
 Best Local Similarity 23.3%; Pred. No. 2.8;
 Matches 45; Conservative 21; Mismatches 65; Indels 62; Gaps 8;
 QY 50 LEHQ-----LRSLAGTYLVNLPFPNPPPLGAEITLPRATVNLVWRSNDRLRLTQ 105
 DB 842 LQHRPVVLFGLAGTAPKLPGPQOEAVTLPNHLQIL-----SIPGWRYSNHTHSQ 896
 QY 106 NYEAYSHLLCYLRLGNLROAATAELRRSLAHF-----CTSLQGLLGSIG 149
 DB 897 N-----LRKHGRGEAQADLRVLLRLYHLYEVGEDPVLSPQVTVNLQVLQALGS 946
 QY 150 VMATLGYPLPQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWLRSAKDFNR--LKK 207
 DB 947 VVAV-----EERSLGT-----WDLMLHRWSWRTGPRHRRGDTTS 982
 QY 208 KMOPPAAS-VTLH 219
 DB 983 PSRPPGGPIITVH 995
 RESULT 7
 NXF1 HUMAN
 ID NXF1 HUMAN STANDARD; PRT; 619 AA.
 AC Q9UBU9; Q9UQL2; Q99799;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated
 DE protein) (mRNA export factor TAP).
 GN NXF1 OR TAP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=99219873; PubMed=10202158;
RA Braun I.C., Rohrbach E., Schmitt C., Izaurralde E.;
RT "TAP binds to the constitutive transport element (CTE) through a novel
RNA-binding motif that is sufficient to promote CTE-dependent RNA
export from the nucleus.";
RL EMBO J. 18:1953-1965(1999).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=99257272; PubMed=10323864;
RA Kang Y., Cullen B.R.;
RT "The human Tap protein is a nuclear mRNA export factor that contains
novel RNA-binding and nucleocytoplasmic transport sequences.";
RL Genes Dev. 13:1126-1139(1999).
[3]
RN SEQUENCE FROM N.A.
RP MEDLINE=99384298; PubMed=10454577;
RA Bear J., Tan W., Zolotukhin A.S., Tabernero C., Hudson E.A.,
RA Felber B.K.;
RT "Identification of novel import and export signals of human TAP, the
protein that binds to the constitutive transport element of the type
D retrovirus mRNA.";
RL Mol. Cell. Biol. 19:6306-6317(1999).
[4]
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RP TISSUE=Lung;
RC Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Oraki K., Hirao M., Ohmori Y., Oka T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE OF 61-619 FROM N.A.
RP TISSUE=Lymphocytes;
RC Yoon D.-W., Lee H., Seol W., DeMaria M., Rosenzweig M., Jung J.U.;
RA "Tap: a novel cellular protein that interacts with tip of herpesvirus
saimiti and induces lymphocyte aggregation.";
RL Immunity 6:571-582(1997).
[7]
RN FUNCTION.
RP PubMed=9660949;
RA Grueter P., Tabernero C., von Kobbe C., Schmitt C., Saavedra C.,
RA Bachi A., Wilk M., Felber B.K., Izaurralde E.;
RT "TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export
from the nucleus.";
RL Mol. Cell 1:649-659(1998).
[8]
RN CHARACTERIZATION.
RP MEDLINE=21282872; PubMed=11259411;
RA Braun I.C., Herold A., Rode M., Conti E., Izaurralde E.;
RT "Overexpression of TAP/p15 heterodimers bypasses nuclear retention and
stimulates nuclear mRNA export.";
RL J. Biol. Chem. 276:20536-20543(2001).
[9]
RN CHARACTERIZATION.
RP PubMed=1066806;
RA Bachi A., Braun I.C., Rodrigues J.P., Pante N., Ribbeck K.,
RA von Kobbe C., Kutay U., Wilk M., Gorlich D., Carmo-Fonseca M.,
RA Izaurralde E.;
RT "The C-terminal domain of TAP interacts with the nuclear pore complex
and promotes export of specific CTE-bearing RNA substrates.";
RL RNA 6:136-158(2000).
[10]
RN MUTAGENESIS.
RP MEDLINE=21151125; PubMed=11256625;
RA Suyama M., Doerks T., Braun I.C., Sattler M., Izaurralde E., Bork P.;
RT "Prediction of structural domains of TAP reveals details of its
interaction with p15 and nucleoporins.";
RL EMBO Rep. 1:53-58(2000).
[11]
RN X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 102-372.
RP MEDLINE=20514125; PubMed=11060011;
RA Likier E., Fernandez E., Izaurralde E., Conti E.;
RT "The structure of the mRNA export factor TAP reveals a cis arrangement
of a non-canonical RNP domain and an LRR domain.";
RL EMBO J. 19:5587-5598(2000).
[12]
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH NXT1, AND X-RAY
RP CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEX WITH NXT1-FG-REPEAT.
RX PubMed=11583626;
RA Fribourg S., Braun I.C., Izaurralde E., Conti E.;
RT "Structural basis for the recognition of a nucleoporin FG repeat by
the NTF2-like domain of the TAP/p15 mRNA nuclear export factor.";
RL Mol. Cell 8:645-656(2001).
[13]
RN STRUCTURE BY NMR OF 551-619, AND MUTAGENESIS OF PHE-617.
RP MEDLINE=21912422; PubMed=11875519;
RA Grant R.P., Hurt E., Neuhaus D., Stewart M.;
RT "Structure of the C-terminal FG-nucleoporin binding domain of
Tap/NXF1.";
RL Nat. Struct. Biol. 9:247-251(2002).
CC -!- FUNCTION: Involved in the nuclear export of mRNA species bearing
retroviral constitutive transport elements (CTE) and in the export
of mRNA from the nucleus to the cytoplasm.
CC -!- SUBUNIT: Interacts with NXT1, NXT2, E1B-AP5, RAE1, the REF
proteins and with several nucleoporins
CC -!- SUBCELLULAR LOCATION: Nuclear; localized predominantly in the
nucleoplasm and at both the nucleoplasmic and cytoplasmic faces of
the nuclear pore complex. Shuttles between the nucleus and the
cytoplasm.
CC -!- TISSUE SPECIFICITY: Expressed ubiquitously.
CC -!- DOMAIN: The minimal CTE binding domain consists of an RNP-
type RNA binding domain (RBD) and leucine-rich repeats.
CC -!- DOMAIN: The nucleoporin binding domain consists of a NTF2-like
domain and a UBA-like domain. The NTF2 domain heterodimerizes with
NXT1 AND NXT2. The formation of NXF1/NXT1 heterodimers is required
for NXF1-mediated nuclear mRNA export. The UBA-like domain
mediates direct interactions with nucleoporin-FG-repeats and is
necessary and sufficient for localization of NXF1 to the nuclear
rim. The conserved loop 594-NWD-596 of the UBA domain has a
critical role in the interaction with nucleoporins.
CC -!- DOMAIN: The leucine-rich repeats and the NTF2-domain are
essential for the export of mRNA from the nucleus.
CC -!- MISCELLANEOUS: The RNA-binding domain is a non-canonical RNP-type
domain.
CC -!- SIMILARITY: BELONGS TO THE NXF FAMILY.
CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).
CC -!- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -!- SIMILARITY: CONTAINS 1 UBA DOMAIN.

CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL: AJ132712; CAA10753.1; -;
DR EMBL: AF112880; AAD39102.1; -;
DR EMBL: AF126246; AAD20016.1; -;
DR EMBL: BC004904; AAH04904.1; -;
DR EMBL: AK027192; -; NOT ANNOTATED_CDS.
DR EMBL: U80073; AAB81111.1; -;

```
DR Genew; HGNC:8071; NXF1.
DR MIM; 602647; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR PDB; 1GO5; 18-OCT-01.
DR PDB; 1JRG; 12-JUL-01.
DR PDB; 1JNS; 23-JUL-01.
DR Pfam; PF00560; LRR; 2.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00446; LRRcap; 1.
DR PROSITE; PS50177; NTF2_DOMAIN; 1.
KW Transport; mRNA transport; Nuclear protein; RNA-binding; Repeat;
KW Leucine-rich repeat; Multigene family; 3D-structure.
FT DOMAIN 119 198
FT REPEAT 266 291
FT REPEAT 292 315
FT REPEAT 316 343
FT REPEAT 344 371
FT REPEAT 386 536
FT DOMAIN 566 608
FT DOMAIN 67 100
FT DOMAIN 83 110
FT DOMAIN 551 561
FT MUTAGEN 306 308
FT MUTAGEN 594 594
FT MUTAGEN 595 595
FT MUTAGEN 617 617
FT CONFLICT 119 119
FT CONFLICT 256 256
SQ SEQUENCE 619 AA; 70182 MW; 338872AADA789FBF CRC64;

Query Match
Best Local Similarity 7.0%; Score 86; DB 1; Length 619;
Matches 61; Conservative 32; Mismatches 89; Indels 58; Gaps 14;

Qy 5 AGDSWGL-----ACLCTVLMHLPALNRTGDP-GFGPSIQKTYDIT---RYLEHQL 54
Db 400 SGRQGLLDAYHDGACCSLIPFQNPAPRSSLAIEYFKDSNNVKLKDPTRLRFLKKHTR 459
Qy 55 RSLATYLYNLGPPNEPPFNP--FRLGAE--TLPRATVN---LEWRSINDBLR-LTQN 106
Db 460 LNVV-AFLNEL--PKTQHDVNSFVVDISAQTSLTLCFSVNGVFKEDYSGKSRDSLRAFTRT 516
Qy 107 Y---EAYSHLLCYLRG--LNRQAATAELRRSLAHFCTSLQGLLGSIAGVMATLGYPLPOP 161
Db 517 FIAPVPSNSGLCIVNDELVEVRNASEIEIQAFV-----MPAP 553
Qy 162 LPTGTEPAWAPGPAHSDFLQK-----MDDFWLLKEIQTLW-----RSKADFNRLKKKQVPP 212
Db 554 TPSSSPVPTLSPEQEMQLQAFSTQSGMNLWSQKCLQDNNWDYTRSAQAPTHLKAKGEIP 613

RESULT 8
MIP1 SCHPO STANDARD; PRT; 1313 AA.
AC P87141.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WD-repeat protein mipl.
GN MIP1 OR SPAC57A7.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=20115869; PubMed=10648609;
RA Shinozaki-Yabana S., Watanabe Y., Yamamoto M.;
RT "Novel WD-repeat protein Mip1p facilitates function of the meiotic
RT regulator Mei2p in fission yeast."
```

```
RL Mol. Cell. Biol. 20:1234-1242(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney J., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Calabret F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., Del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: BINDS TO AND FACILITATES THE FUNCTIONING OF THE MEIOTIC
CC REGULATOR MEI2. MAY ALSO BE INVOLVED IN CONJUGATION BY INTERACTING
CC WITH STE11. ESSENTIAL FOR CELL GROWTH.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB032552; BAA84585.1; -.
CC EMBL; Z95396; CAB08769.1; -.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 5.
CC PRINTS; PR00320; GPROTEINBRPT.
CC PRINTS; PR01547; YEAST176DUF.
CC SMART; SM00320; WD40; 5.
CC PROSITE; PS00678; WD_REPEATS_1; 1.
CC PROSITE; PS00682; WD_REPEATS_2; 2.
CC PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Meiosis; WD repeat; Repeat.
FT REPEAT 986 1029 WD 1.
FT REPEAT 1033 1074 WD 2.
FT REPEAT 1087 1126 WD 3.
FT REPEAT 1130 1170 WD 4.
FT REPEAT 1176 1216 WD 5.
FT REPEAT 1219 1259 WD 6.
FT REPEAT 1268 1308 WD 7.
SQ SEQUENCE 1313 AA; 148533 MW; C71B663B0171E7A4 CRC64;

Query Match
Best Local Similarity 7.0%; Score 86; DB 1; Length 1313;
Matches 58; Conservative 31; Mismatches 84; Indels 88; Gaps 12;

Qy 15 ICTVLMHLPV--PALNR-----TGDPGGPSI-----OKTYDLTRYLHQLS 56
Db 811 LAFLLQHLPALHKASLKSDTNTSVTDKPKHPFVPSVSENKILNRFSLTRSLKGLAUS 870
```

```

QY 57 LAG-----TYLNYL-----GPP-FNEPDPNPPRLGAETLPRATVNL 91
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 871 LAGDRASELLSNGENKPAESNLNHLTSAKVPGPFAFNELEYQ-----SELDMLTSLYLF 926
QY 92 EVWRLNDRRLRLTONYAYSHLLCYLRLG-----NRQAATAELRRSLAHFCTS-----LOG 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 927 DWSRKYFTFPQWRNEDDEPGSICNORLWRNRNREKLIYRTRPLAEYSTNGRWNOQLMT 986
QY 143 LLGSIAGVWATLGVPPLPQPLPOTEPAPGPAHSDFLQKMDDFWLKKE---LQTLWLRSA 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 987 FNVITIA-----PRKLMFHQEPDQLITLGDKDIQVWDWRN 1022
QY 200 KDFNRLKKMQPPAASVTLHL 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 RCLNSFK-----TSASATTV 1038

RESULT 9
PHAC PSEOL STANDARD; PRT; 560 AA.
AC P26496;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Poly(3-hydroxyalkanoate) polymerase 2 (EC 2.3.1.-) (PHA polymerase 2)
DE (PHA synthase 2) (Polyhydroxyalkanoic acid synthase 2).
GN PHAC.
OS Pseudomonas oleovorans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=301;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GP01;
RX MEDLINE=91115830; PubMed=1989978;
RA Huisman G.W., Wonink E., Meima R., Kazemier B., Terpstra P.,
RA Witholt B.;
RT "Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas
RT oleovorans. Identification and sequences of genes and function of the
RT encoded proteins in the synthesis and degradation of PHA.";
RL J. Biol. Chem. 266:2191-2198(1991).
CC -1- FUNCTION: P.OLEOVORANS ACCUMULATES POLY(3-HYDROXYALKANOATES) AFTER
CC GROWTH ON MEDIUM CHAIN LENGTH HYDROCARBONS. LARGE AMOUNTS OF THIS
CC POLYESTER ARE SYNTHESIZED WHEN CELLS ARE GROWN UNDER NITROGEN-
CC LIMITING CONDITIONS. WHEN NITROGEN IS RESUPPLIED IN THE MEDIUM,
CC THE ACCUMULATED PHA IS DEGRADED.
CC -1- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M58445; AAA25934.1; -.
DR PIR; C38604; C38604.
DR InterPro; IPR000073; Abhydrolase.
DR Pfam; PF00561; abhydrolase.1.
KW PHA biosynthesis; Transferase; Acyltransferase.
FT ACT SITE 296 296 POTENTIAL.
SQ SEQUENCE 560 AA; 62631 MW; EACD844FC1616B83 CRC64;

Query Match 7.0%; Score 85.5; DB 1; Length 560;
Best Local Similarity 30.2%; Pred. No. 2.3;
Matches 38; Conservative 16; Mismatches 43; Indels 29; Gaps 7;

QY 77 PRLGAETLPRATVNLVWRSNDRRLRLTONYAYSHLLCYLRLGRLNQAATAELRRSL--A 134
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 5 PAKGTPTLPATSMNVQ-----NAILGLRGR-----DLISTLRNVSQRS-----LRHPLHTA 50

```

```

QY 135 HFCTSLQGLGSIAGVWATLGVPPLPQPLPG-----TEPAWAPGPAHSDFLQKMDDFWLKKE 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 HHLLAGGQLGRV-----ILGDTPLQBNPRDPRSDPTWSQNPVYRRGLQA-----YLAWQ 101
QY 191 LQTLWLW 196
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 102 KQTRLW 107
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 10
HUPK RHOCA STANDARD; PRT; 294 AA.
AC P30797;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase expression/formation protein hupK.
GN HUPK.
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91117833; PubMed=2007559;
RA Xu H.W., Wall J.D.;
RT "Clustering of genes necessary for hydrogen oxidation in Rhodobacter
RT capsulatus.";
RT J. Bacteriol. 173:2401-2405(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33303 / B10;
RX MEDLINE=93268090; PubMed=8497190;
RA Colbeau A., Richard P., Toussaint B., Caballero F.J., Elster C.,
RA Delphin C., Smith R.L., Chabert J., Vignais P.M.;
RT "Organization of the genes necessary for hydrogenase expression in
RT Rhodobacter capsulatus. Sequence analysis and identification of two
RT hyp regulatory mutants.";
RL Mol. Microbiol. 8:15-29(1993).
CC -1- SIMILARITY: BELONGS TO THE HUPK FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M55089; AAA72924.1; -.
DR EMBL; Z15089; CAA78803.1; -.
DR PIR; B38532; B38532.
DR PIR; B38532; B38532.
FT CONFLICT 70 70 A -> R (IN REF. 2).
SQ SEQUENCE 294 AA; 30222 MW; 2669EB70AF35EDA CRC64;

Query Match 6.9%; Score 84.5; DB 1; Length 294;
Best Local Similarity 31.2%; Pred. No. 1.2;
Matches 40; Conservative 3; Mismatches 38; Indels 47; Gaps 6;

QY 76 PPRLGAETLPRATVNLVWRSNDRRLRLTONYAYSHLLCYLRLGRLNQAATAE---LRRS 132
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 36 PPOQVALLPR-----LNLGGAAGHARLAL-----GLPAAPARREILRDH 80
QY 133 LAHFCTSLQGLGSIAGVWATLGVPPLPQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQ 192
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 81 LAKLCLLWPKLJGU-----APQPLP-----EHWAGGA-----ALQ 111
QY 193 TNLWRSK 200
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 112 HNLWGGAK 119
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

```
RESULT 11
M2B2_PIG          STANDARD;          PRT;          995 AA.
AC Q28949;
DT 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epididymis-specific alpha-mannosidase precursor (EC 3.2.1.24)
DE (Mannosidase alpha class 2B member 2) (AMAN).
GN MAN2B2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Epididymis; PubMed=8562059;
RX Okamura N., Tamba M., Liao H.-J., Onoe S., Sugita Y., Dacheux F.,
RA Dacheux J.-L.;
RT "Cloning of complementary DNA encoding a 135-kilodalton protein
RT secreted from porcine corpus epididymis and its identification as an
RT epididymis-specific alpha-mannosidase."
RL Mol. Reprod. Dev. 42:141-148(1995).
CC -!- FUNCTION: CAN DIGEST BOTH P-NITRO-PHENYL-ALPHA-D-MANNOSIDE AND
CC HIGH MANNOSE OLIGOSACCHARIDE (MAN(8)-GLCNAC(2)). MAY BE INVOLVED
CC IN SPERM MATURATION. HAS A POSSIBLE ROLE IN SPECIFIC SPERM-BGG
CC INTERACTION SINCE SPERM SURFACE MANNOSIDASE ACTS LIKE A RECEPTOR
CC FOR MANNOSE-CONTAINING OLIGOSACCHARIDES LOCATED ON THE ZONA
CC PELLUCIDA.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC mannose residues in alpha-D-mannosides.
CC -!- SUBCELLULAR LOCATION: SECRETED. FOUND AT THE SPERM SURFACE AS A
CC 27 kDa FRAGMENT.
CC -!- TISSUE SPECIFICITY: SPECIFIC TO THE CAPUT AND CORPUS OF THE
CC EPIDIDYMIS.
CC -!- PTM: PROCESSED INTO A 27 kDa FRAGMENT LOCALIZED ON THE EQUATORIAL
CC SEGMENT AND THE APICAL RIM OF THE HEAD OF MATURE SPERM.
CC -!- MISCELLANEOUS: OPTIMAL MANNOSIDASE ACTIVITY WAS FOUND AT PH 6.5.
CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D28521; BAA05877.1; ALT INIT.
DR InterPro; IPR000602; Glyco_hydro_38.
DR Pfam; PF01074; Glyco_hydro_38; 2_
KW Hydrolase; Glycosidase; Signal; Glycoprotein.
FT SIGNAL
FT CHAIN 1 21
FT FT 22 995 EPIDIDYMIS-SPECIFIC ALPHA-MANNOSIDASE.
FT FT 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 625 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 657 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 733 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 793 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 793 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 875 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 977 N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT 945 D -> A (IN REF. 1; AA SEQUENCE).
FT CONFLICT 945 945
FT SEQUENCE 995 AA; 114230 MW; 79D21B8CE5AE6FC2 CRC64;
Query Match 6.6%; Score 80.5; DB 1; Length 995;
Best Local Similarity 23.0%; Pred. No. 14;
Matches 45; Conservative 22; Mismatches 62; Indels 67; Gaps 9;
QY 50 LEHQ-----LRSAGYLYNLVGGPPFNEPDNPRLGAEYL-----PRATVLEWRSL 97
DB 827 LQHRPVLFRELGTVQNGFGPRKQEPVTLPPSLHLQILSPGWKYSNHTVHLKNLQKG 886
```

```
QY 98 NDR-----LRLTONYEAYSHLLCYLRGLNRQATAELRRSLAHFCTSLQGLLSI 147
DB 887 HYERRAKADFRVLLRLHLYEAGEH-----QALSRPVT-----LNLSQVLRGL 929
QY 148 AGVMATLGYPLPQPLPQTEPANAPGPAHSDFLQKMDDFWLKELQTLWLRSAKPNR--L 205
DB 930 GSVVAV-----EERSLTGT-----MDVNSLHRWSWKTEDGHHRRGS 965
QY 206 KKKMQPP--AASVTLH 219
DB 966 SRRPLPLRGPNVTIH 981
RESULT 12
VPP3_HUMAN
ID VPP3_HUMAN STANDARD; PRT; 830 AA.
AC Q13488; O75877;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vacuolar proton translocating ATPase 116 kDa subunit A isoform 3 (V-
DE ATPase 116-kDa isoform a3) (Osteoclastic proton pump 116 kDa subunit)
DE (OC-116 kDa) (OC116) (T-cell immune regulator 1) (T cell immune
DE response cDNA7 protein) (TIRC7).
GN TCIRG1 OR ATP6N1C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Osteoclastoma;
RX MEDLINE=96158968; PubMed=8579597;
RA Li Y.P., Chen W., Stashenko P.;
RT "Molecular cloning and characterization of a putative novel human
RT osteoclast-specific 116-kDa vacuolar proton pump subunit.";
RL Biochem. Biophys. Res. Commun. 218:813-821(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RA Utku N., Heinemann T., Bulwin C.-G., Beinke S., Beato F., Randall J.,
RA Busconi L., Delphire E., Robertson E.R., Kojima R., Volk H.D.,
RA Milford E.L., Gullans S.R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RX MEDLINE=99263502; PubMed=10329006;
RA Heinemann T., Bulwin G.C., Randall J., Schnieders B., Sandhoff K.,
RA Volk H.D., Milford E., Gullans S.R.;
RT "Genomic organization of the gene coding for TIRC7, a novel membrane
RT protein essential for T cell activation.";
RL Genomics 57:398-406(1999).
CC -!- FUNCTION: PART OF THE PROTON CHANNEL OF V-ATPASES (BY SIMILARITY).
CC SEEMS TO BE DIRECTLY INVOLVED IN T CELL ACTIVATION.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: THE LONG ISOFORM IS HIGHLY EXPRESSED IN
CC OSTEOCLASTOMAS. THE SHORT FORM IS HIGHLY EXPRESSED IN THYMUS.
CC -!- SIMILARITY: BELONGS TO THE V-ATPASE 116 kDa SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
```

```

DR EMBL: U45285; AAA97878.1; -.
DR EMBL: AF025374; AAC35742.1; -.
DR EMBL: AF033033; AAD31081.2; -.
DR GenBank: HGNC:11647; TCIRG1.
DR MIM: 604592; -.
DR InterPro: IPR002490; V_ATPase_sub116.
DR Pfam: PF01496; V_ATPase_sub_a; 1.
KW Hydrogen ion transport; Transmembrane; Glycoprotein;
KW Alternative splicing.
FT DOMAIN 1 397 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 445 465 POTENTIAL.
FT TRANSMEM 504 524 POTENTIAL.
FT TRANSMEM 538 558 POTENTIAL.
FT TRANSMEM 576 596 POTENTIAL.
FT TRANSMEM 636 656 POTENTIAL.
FT TRANSMEM 771 791 POTENTIAL.
FT CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 503 503 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1 216 MISSING (IN SHORT ISOFORM).
FT CONFLICT 377 377 A -> R (IN REF. 1).
FT CONFLICT 603 603 MISSING (IN REF. 1).
SQ SEQUENCE 830 AA; 92997 MW; 62EAE9A22DC698B CRC64;

Query Match 6.5%; Score 79.5; DB 1; Length 830;
Best Local Similarity 24.3%; Pred. No. 13;
Matches 45; Conservative 25; Mismatches 66; Indels 49; Gaps 9;

Qy 40 IQKTYDLTRYLEHQRLSLAGTYLVNLYGPPFPNPNRLGNATLPRAVNVLEWVRSND 99
Db 61 LEKTP---TFLQEEVRR-AGL---VLPPPKGLRPAPPFR-----D 93

Qy 100 RLRLTQNYEAYSHLLCYLRGLNRQAATLRLSLAHFCTSLQGLLSIAGVM---ATLGY 156
Db 94 LLRIQEETERIAQELRDVRG-NQQLRALQHLQHLQAAVLRQGHFPQLAAATDGSERT 152

Qy 157 PLPQPLPGTEPAWAPGPAHSD----FLQKMDDFWLKXELQTLWRSASK-----DFNRLKK 207
Db 153 PLLQ-----APGPHQDLRVNFVAGVPEPKAPALERLLWRACRGLFIASFLEEQ 203

Qy 208 KNOPP 212
Db 204 PLEHP 208

RESULT 13
YM32_MYCTU STANDARD; PRT; 291 AA.
AC Q10515; Q10516;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RV2232/RV2233.
GN RV2232/RV2233 OR MT2292 OR MTCY427.13/MTCY427.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T.E., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
```

```

RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: BELONGS TO THE CBBY/CBBZ/GPH/YIEH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z70692; CAA94666.1; ALT TERM.
DR EMBL: Z70692; CAA94655.1; ALT_INIT.
DR TIGR; MT2292; -.
DR TubercuList; RV2232; -.
DR InterPro: IPR001454; Hlg_nase/hydrase.
DR Pfam: PF00702; Hydrolase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 291 AA; 30694 MW; 750F090FB154E6E5 CRC64;

Query Match 6.5%; Score 79; DB 1; Length 291;
Best Local Similarity 25.8%; Pred. No. 4;
Matches 47; Conservative 16; Mismatches 65; Indels 54; Gaps 11;

Qy 23 PAVPALNRTGDPGPGPSIQKTYDLTRYLEHQRLSLAGTY---LNYLGGPPFPNPDF----- 74
Db 64 PPAARARRASSPGSPQL-VIFDLDTGTLTDSARGIVSSFRHNLNIGAPVPEGLATHIV 122

Qy 75 NPPLRGAETLP-----RATNVLEWVRSNDRLRLTQNYEAYSHLLCYLR- 118
Db 123 GPPM--HETLRAMGLGESAEFAIVAYRADYSARGW-----AMNSLFDGIGPLDLRT 173

Qy 119 -GLNRQAATAE---LRSLAHF-----CTSLQGLLSIAGVMATLGYPLPQ--P 161
Db 174 AGVRLAVATSKAEPTRARRILRHFGIEQHFVFIAGASTDGSRGSKVDVLA---HALAQLRP 230

Qy 162 LP 163
Db 231 LP 232

RESULT 14
BAT3_HUMAN STANDARD; PRT; 1132 AA.
ID BAT3_HUMAN
AC P46379;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Large proline-rich protein BAT3 (HLA-B-associated transcript 3).
GN BAT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=T-cell;
RX MEDLINE=90192810; PubMed=2156268;
RA Banerji J., Sands J., Strominger J.L., Spies T.;
RT "A gene pair from the human major histocompatibility complex encodes
```


large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain.";
Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
-!- FUNCTION: UNKNOWN.
-!- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M33519; AAA35587.1; -.
EMBL; M33521; AAA35588.1; -.
EMBL; M33520; AAA35588.1; JOINED.
HSP; P02248; IUBI.
Genew; HGNC:13919; BAT3.
MIM; 142590; -.
InterPro; IPR000626; Ubiquitin.
Pfam; PF00240; ubiquitin; 1.
SMART; SM00213; UBQ; 1.
PROSITE; PS00299; UBIQUITIN_1; 1.
PROSITE; PS50053; UBIQUITIN_2; 1.
KW Repeat.
FT DOMAIN 17 77 UBIQUITIN-LIKE.
FT DOMAIN 202 207 POLY-PRO.
FT DOMAIN 242 636 4 X 29 AA APPROXIMATE REPEATS.
FT REPEAT 242 270 1.
FT REPEAT 415 443 2.
FT REPEAT 474 602 3.
FT REPEAT 608 636 4.
FT DOMAIN 657 670 POLY-PRO.
FT SEQUENCE 1132 AA; 119504 MW; 528CA8A78C38DD18 CRC64;
SQ
Query Match 6.5%; Score 79; DB 1; Length 1132;
Best Local Similarity 26.1%; Pred. No. 22;
Matches 46; Conservative 17; Mismatches 67; Indels 46; Gaps 11;
QY 23 PAVPALNRGTGPGPSIQKTYDLYLEHQLSLAGTYLNYLGPFPNEPDFNPPRLGAE 82
Db 261 PA-PETNAPNHSPARYBVVLOELQR-LRSRLQPFQRYEVLGAAAT-TDYNNHEGRE 317
QY 83 TLPRATVNLVWRSNDRLRLTON-YEAYSHLLCYLRGLNRQAAT----AELRRSLAHFC 137
Db 318 EQDR-LINL-----VGESLRLLGNTFVALSDLRCLN-----ACTPRHLHVVRPKSHYT 365
QY 138 TSLOGLGSIAGMATLGYPL-----POPLPQTE-PAWAPGPAHS 176
Db 366 TPM-----VLQQAALPIQINVTGTTMTGNGTRPPTPNAEAPPPGPGQASS 412
RESULT 15
ID ALK HUMAN STANDARD; PRT; 1620 AA.
AC Q9UM73; Q9Y4K6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ALK tyrosine kinase receptor precursor (SC 2.7.1.112) (Anaplastic lymphoma kinase) (CD246 antigen).
GN ALK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND RP GLYCOSYLATION.
RX MEDLINE=97316779; PubMed=9174053;
RA Morris S.W., Naeve C.W., Mathew P., James P.L., Kirstein M.N., Cui X., Witte D.P.;

"ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK).";
Oncogene 14:2175-2188(1997).
[2]
RN ERRATUM.
RP Morris S.W., Naeve C.W., Mathew P., James P.L., Kirstein M.N., Cui X., Witte D.P.;
RA Oncogene 15:2883-2883(1997).
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=97178863; PubMed=9053841;
RA Iwahara T., Fujimoto J., Wen D., Cupples R., Bucay N., Arakawa T., Mori S., Razukin B., Yamamoto T.;
RT "Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system.";
Oncogene 14:439-449(1997).
[4]
RN PARTIAL SEQUENCE FROM N.A., AND CHROMOSOMAL TRANSLOCATION.
RX MEDLINE=94167588; PubMed=812112;
RA Morris S.W., Kirstein M.N., Valentine M.B., Dittmer K.G., Shapiro D.N., Saltman D.L., Look A.T.;
RT "Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma.";
RL Science 263:1281-1284(1994).
CC -!- FUNCTION: Orphan receptor with a tyrosine-protein kinase activity. Appears to play an important role in the normal development and function of the nervous system.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in brain and CNS. Also expressed in the small intestine and testis, but not in normal lymphoid cells.
CC -!- PTM: N-glycosylated.
CC -!- DISEASE: A FORM OF NON-HODGKIN'S LYMPHOMA IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(2;5) (P23;Q35) THAT INVOLVES NPM1 AND ALK.
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN RECEPTOR SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 MAM DOMAINS.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chromocancer/Genes/ALK.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U62540; AAB71619.1; -.
DR EMBL; U66559; AAC51104.1; -.
DR HSSP; P08631; 1AD5.
DR Genew; HGNC:427; ALK.
DR MIM; 105590; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR002011; RTKinaseII.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS50060; MAM_2; 2.
DR PROSITE; PS01209; LDLRA_1; FALSE_NEG.
DR PROSITE; PS50068; LDLRA_2; FALSE_NEG.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
KW Transferase; Tyrosine-protein kinase; Transmembrane; ATP-binding;
KW Phosphorylation; Receptor; Glycoprotein; Repeat; Signal;
KW Proto-oncogene; Chromosomal translocation.
FT SIGNAL 1 18
FT CHAIN 19 1620
FT DOMAIN 19 1038
FT TRANSF 1039 1059
FT DOMAIN 1060 1620
FT DOMAIN 264 427
FT DOMAIN 437 473
FT DOMAIN 478 636
FT DOMAIN 1116 1392
FT DOMAIN 816 940
FT NP_BIND 1122 1130
FT BINDING 1150 1150
FT ACT_SITE 1249 1249
FT MOD_RES 1282 1282
FT CARBOHYD 169 169
FT CARBOHYD 244 244
FT CARBOHYD 285 285
FT CARBOHYD 324 324
FT CARBOHYD 411 411
FT CARBOHYD 424 424
FT CARBOHYD 445 445
FT CARBOHYD 563 563
FT CARBOHYD 571 571
FT CARBOHYD 627 627
FT CARBOHYD 709 709
FT CARBOHYD 808 808
FT CARBOHYD 863 863
FT CARBOHYD 864 864
FT CARBOHYD 886 886
FT CARBOHYD 986 986
FT CONFLICT 36 36
FT CONFLICT 1491 1491
FT CONFLICT 1529 1529
SQ SEQUENCE 1620 AA; 176417 MW; A62604B242961E1E CRC64;

Query Match 6.5%; Score 79; DB 1; Length 1620;
Best Local Similarity 23.1%; Pred. No. 35;
Matches 54; Conservative 15; Mismatches 77; Indels 88; Gaps 9;

Qy 18 VLWHLPAVPAL-----NRTGDPGPGSIQKTYDITRYLHQLSLAGTYLNLGP 67
Db 6 LLWLLPLLSTAAVSGMGQTGORAGSPAAGSPLPQPREPLS-YSRLQKSLA----- 55
Qy 68 PFNEPDPNPRLGAEATLPRAATVNLVWRSLNDRRLTQNYEAYSH-LLCYLRLGLNROAAT 126
Db 56 ----VDFVVPSL-----FRVYARDLLPPSSSELKAGR 84
Qy 127 AELRRSLAHFTSLQGLLSTAGVNVATLGYPLQPLP----- 163
Db 85 PEARGSLALDCAPLRLLLGPAFGVSWTAGSPAPAEARTLSRLVKGGSVKLRARAKQLVLE 144
Qy 164 -GTEP-----AWAPGPAHSDFLQKDDFWLLKELQTLWRSKADFNLKXKMQP 211
Db 145 LGEEAILGECVGPGEAAVGLLQ-----FNLSELFSWIIRQE--GRUKIRLMP 191

Search completed: January 27, 2003, 15:42:00
Job time : 11 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:23 ; Search time 26.5 Seconds
(without alignments)
1749.457 Million cell updates/sec

Title: US-09-931-704-5
Perfect score: 1224
Sequence: 1 MDLRAGDSWGMCLCTVLM.....KKMQPPAASVTLHLHAHGF 225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1224	100.0	225	11	Q9QZM3 mus musculus
2	1193	97.5	225	4	Q9UBD9 Q9ubd9 homo sapien
3	144.5	11.8	215	13	Q9PUJ2 Q9puj2 plethodon j
4	144.5	11.8	215	13	Q9PUJ1 Q9puj1 plethodon j
5	144.5	11.8	215	13	Q9PUJ0 Q9puj0 plethodon j
6	142.5	11.6	215	13	Q9PUJ9 Q9puj9 streptomyc
7	93.5	7.6	640	16	O54153 O54153 streptomyc
8	93	7.6	727	11	O88841 O88841 mus musculu
9	92.5	7.6	200	16	Q8YBB1 Q8ybb1 bruceella me
10	92	7.5	332	10	Q9NAU1 Q9naul arabidopsis
11	91.5	7.5	455	11	Q9CWV7 Q9cww7 mus musculu
12	91	7.4	423	11	Q9JHE4 Q9jhe4 mus musculu
13	90	7.4	733	16	Q9HTE5 Q9htc5 pseudomonas
14	89.5	7.3	530	3	Q8XOE9 Q8xoe9 neurospora
15	89	7.3	423	11	Q9DBV6 Q9dbv6 mus musculu
16	89	7.3	733	16	Q91664 Q91664 pseudomonas

17	88	7.2	222	2	Q9L584
18	86.5	7.1	318	4	Q96LS2
19	85.5	7.0	389	17	Q9YA71 Q9ya71 aeropyrum p
20	85.5	7.0	1154	11	Q9Z1R2 Q9z1r2 mus musculu
21	85	6.9	476	11	Q8R363 Q8r363 mus musculu
22	85	6.9	520	16	Q9RT14 Q9rt14 deinococcus
23	85	6.9	585	5	Q9VNR6 Q9vnr6 drosophila
24	84	6.9	741	16	Q9HZM0 Q9hzm0 pseudomonas
25	83	6.8	272	16	Q86721 Q86721 streptomyc
26	83	6.8	396	2	Q9AF00 Q9af00 frankia sp.
27	83	6.8	955	11	O88287 O88287 mus musculu
28	83	6.8	1561	11	O88286 O88286 mus musculu
29	82.5	6.7	1098	11	Q9WTN8 Q9wtm8 rattus norv
30	82.5	6.7	5990	2	Q9LBP6 Q9lbp6 mycobacteri
31	81.5	6.7	209	2	Q9L593 Q9l593 streptococc
32	81.5	6.7	256	2	Q9L590 Q9l590 streptococc
33	81.5	6.7	294	11	Q9CPZ1 Q9cpz1 mus musculu
34	81	6.6	227	2	Q9KGS0 Q9kgs0 streptococc
35	81	6.6	237	2	Q9L592 Q9l592 streptococc
36	81	6.6	256	2	Q9L595 Q9l595 streptococc
37	81	6.6	257	2	Q9L594 Q9l594 streptococc
38	81	6.6	395	2	Q9LAY9 Q9lay9 streptococc
39	81	6.6	1277	3	Q9C2D2 Q9c2d2 neurospora
40	80.5	6.6	373	2	Q9ZGCO Q9zgc0 streptomyc
41	80.5	6.6	379	2	Q9ZGH4 Q9zgh4 streptomyc
42	80	6.5	294	4	Q9NXX5 Q9nxx5 homo sapien
43	80	6.5	382	16	Q8YSG4 Q8ysg4 anabena sp
44	80	6.5	873	10	Q93Y18 Q93y18 corylus ave
45	80	6.5	876	4	Q9P210 Q9p210 homo sapien

ALIGNMENTS

RESULT 1

Q9QZM3	ID	Q9QZM3	PRELIMINARY;	PRT;	225 AA.
AC	Q9QZM3;				
DT	01-MAY-2000	(T-EMBLrel. 13, Created)			
DT	01-MAY-2000	(T-EMBLrel. 13, Last sequence update)			
DT	01-JUN-2001	(T-EMBLrel. 17, Last annotation update)			
DE	Neurotrophin-1/B-cell stimulating factor-3.				
GN	BSF3.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99432254; PubMed=10500198;				
RA	Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,				
RA	Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,				
RA	Simonet W.S., Boone T., Chang M.-S.;				
RT	"Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the				
RT	IL-6 family."				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).				
DR	EMBL; AF176913; AAF00993.1; ..				
DR	MED; MGI:1930088; Bsf3.				
SQ	SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;				

Query Match	100.0%;	Score 1224;	DB 11;	Length 225;
Best Local Similarity	100.0%;	Pred. No. 3.3e-106;		
Matches 225;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MDLRAGDSWGMCLCTVLMHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT	60	
Db	1	MDLRAGDSWGMCLCTVLMHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT	60	
QY	61	YLNVLGPPFPNPPNPRCLGAETLPRATNVLEWRSNDRLRLTONVEAYSHLLCYLRGL	120	
Db	61	YLNVLGPPFPNPPNPRCLGAETLPRATNVLEWRSNDRLRLTONVEAYSHLLCYLRGL	120	
QY	121	NROQATAELRSLAHFCTSLQGLLGIAGVWATLGYPLPQLPCTEPAPWAPGPAHSDFLQ	180	

```
||||| 121 NROAATAELRSIAHFTCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWAPGAHSDFLQ 180
Db
Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKOPPAASVTLHLEAHGF 225
|||||
Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKOPPAASVTLHLEAHGF 225
|||||

RESULT 2
Q9UBD9 PRELIMINARY; PRT; 225 AA.
AC Q9UBD9;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like
cytokine) (Similar to cardiotrophin-like cytokine,
neurotrophin-1/B-cell stimulating factor-3).
GN CLC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,
Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the
IL-6 family.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99382254; PubMed=10448081;
RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,
Ruben S., Alderson R.F.;
RT "Computational EST database analysis identifies a novel member of the
neurotrophic cytokine family.";
RN [3]
RP SEQUENCE FROM N.A.
RX Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF176912; AAF00992.1; -
DR EMBL, AF172854; AAD54284.1; -
DR EMBL, AF176911; AAF00991.1; -
DR EMBL, AY049779; AAL15436.1; -
DR EMBL, BC012939; AAH12939.1; -
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Query Match 97.5%; Score 1193; DB 4; Length 225;
Best Local Similarity 96.3%; Pred. No. 2.6e-103;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGFSIQKTYDLTRYLHQLRSLAGT 60
Db 1 MDLRAGDSWGLACLTIVLWHLPAVPALNRTGDPGPGFSIQKTYDLTRYLHQLRSLAGT 60
Qy 61 YLNYLGPPFPNEPDPNPRPGAETLPRATVNLVWRSLNDRLLRNTQNYEAYSHLLCYLRL 120
Db 61 YLNYLGPPFPNEPDPNPRPGAETLPRATVNLVWRSLNDRLLRNTQNYEAYSHLLCYLRL 120
Qy 121 NROAATAELRSIAHFTCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWAPGAHSDFLQ 180
Db 121 NROAATAELRSIAHFTCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWAPGAHSDFLQ 180
Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKOPPAASVTLHLEAHGF 225
|||||
```

```
||||| 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKOPPAASVTLHLEAHGF 225
Db
Qy 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMKOPPAASVTLHLEAHGF 225
|||||

RESULT 3
Q9PUJ2 PRELIMINARY; PRT; 215 AA.
AC Q9PUJ2;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE Receptivity factor isoform 1 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;
Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
salamander.";
RN [2]
RP SEQUENCE 215 AA; 24138 MW; B1906BB666335738 CRC64;

Query Match 11.8%; Score 144.5; DB 13; Length 215;
Best Local Similarity 25.6%; Pred. No. 1.3e-05;
Matches 41; Conservative 31; Mismatches 81; Indels 7; Gaps 3;

Qy 56 SLACTYNYLGPFPNEPDPNPRPGAETLPRATVNLVWRSLNDRLLRNTQNYEAYSHLLC 115
Db 56 SLACTYNYLGPFPNEPDPNPRPGAETLPRATVNLVWRSLNDRLLRNTQNYEAYSHLLC 115
Qy 55 SLLPTYSFQAGAPISDPDYQLPHIKVANLPAAMDYDTFMKQTDTRLNNNYFYSAYVE 114
Db 55 SLLPTYSFQAGAPISDPDYQLPHIKVANLPAAMDYDTFMKQTDTRLNNNYFYSAYVE 114
Qy 116 YLR-GLNRQ----AATAELRRSLAHFTCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWA 170
Db 116 YLR-GLNRQ----AATAELRRSLAHFTCTSLQGLGSIAGVMTLGYPLPQPLGTEPAWA 170
Qy 115 FLKEAMTEQEDLNPAELSLKAKFEAMANSNTLSKISDINTQMGMSVTITLP--KPLVV 172
Db 115 FLKEAMTEQEDLNPAELSLKAKFEAMANSNTLSKISDINTQMGMSVTITLP--KPLVV 172
Qy 171 PGPAHSDFLQMDDFWLLKELQTLWLRSAKDFNRLKKMKQ 210
Db 171 PGPAHSDFLQMDDFWLLKELQTLWLRSAKDFNRLKKMKQ 210
Qy 173 PFECSAVPRKKLGGVCKEYKERVLLTKRDFEFLAKKYQ 212
Db 173 PFECSAVPRKKLGGVCKEYKERVLLTKRDFEFLAKKYQ 212

RESULT 4
Q9PUJ1 PRELIMINARY; PRT; 215 AA.
AC Q9PUJ1;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE Receptivity factor isoform 2 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;
Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "Proteinaceous pheromone affecting female receptivity in a terrestrial
salamander.";
RN [2]
RP SEQUENCE 215 AA; 24080 MW; B341B8B7B4E28438 CRC64;
```


RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2).";
 DR Nature 417:141-147(2002).
 DR EMBL; AL021409; CAA16181.1; -.
 DR HSSP; P12998; 1BS0.
 DR InterPro; IPR004839; AminoTransf1/2.
 DR InterPro; IPR003880; Pnante. attach.
 DR Pfam; PF00155; aminotran_1_2; 1.
 DR Pfam; PF00550; pp-binding; 2.
 DR PROSITE; PS00075; ACP_DOMAIN; 1.
 KW Phosphopantetheine.
 SQ SEQUENCE 640 AA; E619FBB5B55A613F CRC64;
 Query Match 7.6%; Score 93.5; DB 16; Length 640;
 Best Local Similarity 28.3%; Pred. NO. 2.9;
 Matches 52; Conservative 12; Mismatches 63; Indels 57; Gaps 9;
 QY 22 LPAPVAPALNRGTGPGPSIQKTYDTRYLEHQL---RSLAGTYLVNLYGPPFNEPDENPPR 78
 DB 100 LPAPVAP-EGTGDPTVEAVVAAMRQTQVQHQLGLDADLEG-----E 140
 QY 79 LGAETLPATVNLVYMRSLNDRRLTQNYEAYSHLLCYLRGLNRQAATAEALRRSLAHFCT 138
 DB 141 LGVDSVLTSSVVAEA---TERLGLT-----GAAPDAAGATTTLRALA---D 179
 QY 139 SLOGLLGSIAGVATLGYPLPGLGTEPAWAPGPAHSDFLQKMDDFLLKELQTLWLR 198
 DB 180 ALRGL-----VAAAGPTAVPEAPATGAA-APAPGSGNAP-----APGADGWDHRS 225
 QY 199 AKDF 202
 DB 226 MKDF 229
 RESULT 8
 O88841 PRELIMINARY; PRT; 727 AA.
 AC O88841;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Faciogenital dysplasia protein 2.
 GN FGD2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93389726; PubMed=10458911;
 RA Pastaris N.G.; Gorski J.L.;
 RT "Isolation, characterization, and mapping of the mouse and human fgd2
 RT genes, faciogenital dysplasia (FGD1; aarskog syndrome) gene
 RT homologues.";
 RL Genomics 60:57-66(1999).
 DR EMBL; AF017368; AAC35430.1; -.
 DR MGB; MGI:1347084; FGD2.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; RhoGEF.
 DR InterPro; IPR000306; Znf_FYVE.
 DR Pfam; PF01363; FYVE; 1.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhoGEF; 1.
 DR SMART; SM00064; FYVE; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhoGEF; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR SEQUENCE 727 AA; S2039 MW; 5CD0FBFB905C5FC5 CRC64;
 Query Match 7.6%; Score 93; DB 11; Length 727;
 Best Local Similarity 24.8%; Pred. NO. 3.9;
 Matches 61; Conservative 32; Mismatches 93; Indels 60; Gaps 11;

QY 21 HLPAPVAPALNRGTGPGPSIQKTYDTRYLEHQL-----RSLAGT 60
 DB 13 NLVAVFENNRT--PGEAPGSHSLDQPHIPEHQLSLSPPEWAPPVKEALKSEPRPVSR 70
 QY 61 YLNYLGGPPFN-----EPDFNPPLGATLPRATVNLVWRSLNDRRLTQNYEAYSH 112
 DB 71 YLSSLNKXLSGAWRRSCQPGVSP--GPE-----TOPEBEKRVVRELLETEQAYVARLH 122
 QY 113 LL--CYLRGLNRQAATAEALRRSLAHFCTSLQGLLSIAGVMA-TLGYPLPGLPGLTEPAW 169
 DB 123 LLDQVFFQELLREAG-----RSKAFBEDVVKLIFSNLISSIRFHAQFPLP-LORRVDWM 176
 QY 170 APCPAHSDFLQKMDDFWLL-----KEIQTWLRSAKDFNRLKKMQPPAASVT 217
 DB 177 AATPRIGDVIQKLAPFLKMYSEYVKNFERAAELLATWMDKS-QPFQEVTVTRIQSEASS 235
 QY 218 LHLEAH 223
 DB 236 LTLQHH 241
 RESULT 9
 O8YBB1 PRELIMINARY; PRT; 200 AA.
 AC O8YBB1;
 DT 01-WAR-2002 (TRENBLrel. 20, Created)
 DT 01-WAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Hypothetical protein BMEI10989.
 GN BMEI10989.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G.; Kapatral V.; Redkar R.J.; Patra G.; Mijer C.; Los T.;
 RA Ivanova N.; Anderson I.; Bhattacharya A.; Lykidis A.; Reznik G.;
 RA Jablonski L.; Larsen N.; D'Souza M.; Bernal A.; Mazur M.; Goltzman E.;
 RA Selkov E.; Elzer P.H.; Hagius S.; O'Callaghan D.; Letesson J.-J.;
 RA Haselkorn R.; Kyripides N.; Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009732; AAL54231.1; -.
 DR InterPro; IPR00734; Lipase.
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 200 AA; 21513 MW; B4A8596707B6717 CRC64;

Query Match 7.6%; Score 92.5; DB 16; Length 200;
 Best Local Similarity 23.9%; Pred. No. 0.82;
 Matches 55; Conservative 21; Mismatches 89; Indels 65; Gaps 11;
 QY 10 GMLACLTVLVHLPAVAPALNRGTGPGPSIQKTYDTRYLEHQLRSLAGTYLVNLYGPPF 69
 DB 2 GLLAGAALV---LPSLPAEAKTQQAAMPFNATSPHQADVYL---LRGFADIFSTGI--- 51
 QY 70 NEPDENPPRLCAETLPATVNLV-----WRLNDRRLTQNYEAYSHLLCYLRGLNRQA 124
 DB 52 -----DEIGAE-LQAAGVNAHQHAAWLVNLNRIVADQKNGHLPVVLIGHSLGANA 103
 QY 125 A---TAEALRRSLAHFCTSLQGLLSIAGVMA-TLGYPLPGLPGLT-----EPAW-- 169
 DB 104 AIYIAEELERR-----GIAVDYMATFAATGPDPLPGNVVRVNVNFFYKQHGWL 151
 QY 170 --APGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNRLK-KMQPPAASV 216
 DB 152 PLVPGPRFHGHLENRD-----FSNAKOVGHFNIEKQRPLOAEV 189

SQ SEQUENCE 423 AA; 48984 MW; FD4AA2BA1D983E46 CRC64;

Query Match 7.4%; Score 91; DB 11; Length 423;
Best Local Similarity 22.2%; Pred. No. 3;
Matches 53; Conservative 23; Mismatches 65; Indels 98; Gaps 11;

QY 44 YDLTRYLHQLRSLAGTYLNY---LGPP-----FNEP----- 72
DB 204 YDPSSNAHYLRNLLFFDLGYDSSLDPASPRVQEHLEVERFHLVLLQYFDESIVLLR 263

QY 73 -----DNPPRLGAEATLPRATVNVWRSNDRLRLTONYERAYSHL 113
DB 264 ELLCWDLEBVLVFKLNARSDSPVRLSGELYRRAT-----AMNLLD--VRLYHFNFASFWR 317

QY 114 LCVLRGLNRQA-ATAELRRS---LAHFCTSLQGLGSLA-----GVMATLGY 157
DB 318 KVEAFGRMRAREVNLQANRHEMRHICIDGQAVGAEAIRDSAMQWPQPLGKSLILGN 377

QY 158 LPQPLPGTEPAWAPGPAHSDFLQKM-----DDFWLLKELQTLWRSKADFNR 204
DB 378 LKKSII-----GPQHEQLCRMLTPETQVYLSLDGANLWTK-----LWKFLRDLR 422

RESULT 13

Q9HTCS ID Q9HTCS PRELIMINARY; PRT; 733 AA.

AC Q9HTCS;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE Hypothetical protein PA5441.

GN PA5441.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P., Hickey M.J., Brinkman F.S.D., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen."

RT Nature 406:959-964 (2000).

DR EMBL; AE004957; AAG08826.1; --

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 733 AA; 80036 MW; 83C2DC10D6781A13 CRC64;

Query Match 7.4%; Score 90; DB 16; Length 733;
Best Local Similarity 23.9%; Pred. No. 7.4;
Matches 48; Conservative 24; Mismatches 81; Indels 48; Gaps 8;

QY 23 PAVPALNRTGDPGPGPSIQTKYDLYRLEHQLRSLAGTYLNYLGPPFNEPDPNPRLGAE 82
DB 102 PALPAASAEETPPAPAPSPALAEAR-----OMGAE 132

QY 83 TLPRATVNVWRSNDRLRLTONYERAYSHLLCYLRGLNRQAELR-RSLAHFCTSLQ 141
DB 133 ALPEKIAGAEFEGEGSCRSNDQSA-----LAFLRQV-RDAGLGEATYKALAN---SRL 184

QY 142 GLLGSIAGVMATLGYPLPOPLPGTEPAWAPGPAHSDFLQKDDFWLLKELQTLWRSKAD 201
DB 185 DILGACGWEQEELGGVLAQ---GVES--AAGKAFATYLEAANFYSGR-----FDEAEQG 234

QY 202 FNRLKKKQPPAASVTLHLEA 222

DB 235 FVALQDVQPWLKETALYLQA 255

RESULT 14

Q8X0E9 ID Q8X0E9 PRELIMINARY; PRT; 530 AA.

AC Q8X0E9;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical 59.4 kDa protein.

GN B1486.080.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R., Nyakatura G., Mewes H.W., Mannhaupt G.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL670007; CAD21310.1; --

DR InterPro; IPR001810; F-box.

DR PROSITE; PS0181; FBOX; 1.

KW Hypothetical protein.

SQ SEQUENCE 530 AA; 59352 MW; 70382EE15F71BB9D CRC64;

Query Match 7.3%; Score 89.5; DB 3; Length 530;
Best Local Similarity 23.6%; Pred. No. 5.5;
Matches 39; Conservative 24; Mismatches 65; Indels 37; Gaps 6;

QY 64 YLGPPFNEPDPNPRLGAEATLPRATVNVWRSNDRLRLTONYERAYSHLLCYLRGLNR 122
DB 236 FLVPPYLEPDLPLVLAGKKLHLVLDADLATHRDMELQYLLFKFLALTPNLTLWR---- 291

QY 123 QAATAELRRSLAHFCTSLQGLGSIAGVMATLGYPLPOPLPGT-----EPAWAFGP 173
DB 292 ----VNFRSISHY-----GKGAALLKWLASSLRDCTWSSPGLINADPSRLPPP 336

QY 174 AHSDFLQKDDFWLLKELQTLWRSKADFNRLLKKMQPPAASVTL 218
DB 337 VEPDHLQELD---IGOLDV---SANTLYRLENFKFSSTLKAISL 373

RESULT 15

Q9D8V6 ID Q9D8V6 PRELIMINARY; PRT; 423 AA.

AC Q9D8V6;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Cerebroside (3'-phosphoadenylylsulfate:galactosylceramide 3')

DE sulfotransferase.

GN GCST OR CST.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=PANCREAS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK007645; BAB25160.1; --
DR MGD; MGI:1858277; Gcst.
SQ SEQUENCE 423 AA; 48824 MW; 045303CFECE150A3 CRC64;

Query Match 7.3%; Score 89; DB 11; Length 423;
Best Local Similarity 23.1%; Pred. No. 4.6;
Matches 55; Conservative 21; Mismatches 66; Indels 96; Gaps 11;
Qy 44 YDLTRYLEHQRLSLAGTYINY---LGPP-----FNEP----- 72
Db 204 YDPSSYNAHYLRNLLFFDLGYDSSLDPASPRVQEHILEVERRFHLVLLQEFDESLVLLQ 263
Qy 73 -----DFNPPRLGAETLPRATVNLVWRSLNDRRLTLTONYEAYSHL 113
Db 264 ELLCWDLKDVLVFKLNARRDSPVRLSGELYRRAT-----AWNLLD--VRLYRHFNASFWR 317
Qy 114 LCYLRLGNLNRQA-ATAELRRS---LAHFCTSLQGLLSIA-----GVNATLGYP 157
Db 318 KVEAFGRERMAREVLAELRQANEHMRHICIDGGQAVGAEAIQDSAMQPWQPLGKSLGYN 377
Qy 158 LPQPLPGTEPAWAPGAHSDFLQKMDDFWLLKELQ-----TWLWRSKDFNR 204
Db 378 LKKSII-----GQHEQLCRGM---LTPFIQVLSDLGANLWVTKLWKFLRDFLR 422

Search completed: January 27, 2003, 15:44:09
Job time : 28.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:40:17 ; Search time 33 Seconds
(without alignments)
908.527 Million cell updates/sec

US-09-931-704-5

Title:
Perfect score: 1224

Sequence: 1 MDLRAGDSWGLACLTVM.....KKMQPPAASVTLHLEAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID22/gcgcdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SID22/gcgcdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SID22/gcgcdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SID22/gcgcdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SID22/gcgcdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SID22/gcgcdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SID22/gcgcdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SID22/gcgcdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SID22/gcgcdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SID22/gcgcdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SID22/gcgcdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SID22/gcgcdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SID22/gcgcdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SID22/gcgcdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SID22/gcgcdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SID22/gcgcdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SID22/gcgcdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SID22/gcgcdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SID22/gcgcdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SID22/gcgcdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SID22/gcgcdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SID22/gcgcdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SID22/gcgcdata/geneseq/geneseq-embl/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	100.0	225	AAW29716	Mouse neurotrophic
2	1224	100.0	225	AAW56142	Amino acid sequenc
3	1224	100.0	225	AA1987814	Murine NNT-1 prote
4	1224	100.0	225	AAU78177	Mouse novel neurot
5	1193	97.5	225	AAW29715	Human neurotrophic
6	1193	97.5	225	AAW56141	Amino acid sequenc
7	1193	97.5	225	AAW94466	Human cardiotoxoph
8	1193	97.5	225	AA1987813	Human NNT-1 protei
9	1193	97.5	225	AAG63543	Amino acid sequenc
10	1193	97.5	225	AAU78176	Human novel neurot

11	1193	97.5	253	22	AAW25831	Human protein sequ
12	1181	96.5	223	22	AAE00828	Human cardiotoxoph
13	1171	95.7	321	22	AB11896	Human protein SEQ
14	1171	95.7	321	22	AAW79399	Mouse interleukin-
15	1167	95.3	215	21	AA19587	Human protein SEQ
16	1164	95.1	260	22	AAW78415	Human interleukin-
17	1136	92.8	215	21	AB19586	Peptide #7823 enco
18	852	69.6	164	22	AB140317	Protein #6715 enco
19	852	69.6	164	22	AB247116	Human brain expres
20	852	69.6	164	22	AAW61118	Human bone marrow
21	852	69.6	164	22	AAW73827	Peptide #6549 enco
22	852	69.6	164	22	AAW20115	Peptide #8049 enco
23	852	69.6	164	22	AAW34012	Human peptide enco
24	852	69.6	164	23	ABG43716	Human DNAX interle
25	156.5	12.8	208	20	AAV09197	Human DNAX interle
26	154	12.6	208	20	AAV09196	Mouse cardiac hype
27	110.5	9.0	203	16	AAW83965	Human cardiotoxoph
28	110.5	9.0	203	17	AAW88204	Murine cardiotoxoph
29	110.5	9.0	203	18	AAW29237	Human cytokine Zal
30	92.5	7.6	243	22	AAU09153	Human interleukin
31	92.5	7.6	243	22	AAW20277	Human helical prot
32	92.5	7.6	243	23	AAU76375	Human helical prot
33	92	7.5	332	21	AAW22132	Arabidopsis thalia
34	92	7.5	332	21	AAW40321	Human cardiac hype
35	91.5	7.5	201	16	AAW83967	Human cardiac hype
36	91.5	7.5	201	18	AAW29238	Human tumour-associ
37	91.5	7.5	201	20	AAV06490	Human protein PRO8
38	91.5	7.5	201	21	AAW27662	Human cardiotoxoph
39	91.5	7.5	201	21	AAW13004	Amino acid sequenc
40	91.5	7.5	201	21	AAV3697	Human cardiotoxoph
41	91.5	7.5	201	21	AAW87818	Human cardiotoxoph
42	91.5	7.5	201	22	AAW50994	Human PRO82 prote
43	90	7.4	195	14	AAW34432	Sequence of growth
44	90	7.4	195	20	AAW83337	Chicken ciliary ne
45	87	7.1	242	22	AAW20275	Human interleukin

ALIGNMENTS

RESULT 1
AAW29716
ID AAW29716 standard; Protein; 225 AA.
XX
AC AAW29716;
XX
DT 09-NOV-1998 (first entry)
XX
DE Mouse neurotrophic factor NNT-1.
XX
KW NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW peripheral neuropathy; dystrophy; neural retina degeneration;
KW common variable immunodeficiency; CVID; selective IgA deficiency;
KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
therapy.
XX
OS Mus sp.
XX
FH Key
FT Peptide
FT Protein
FT W09833922-A1.
XX
PD 06-AUG-1998.
XX
PF 02-FEB-1998; 98WO-US02363.
XX
PR 30-JAN-1998; 98US-0016534.

PR 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 PA Chang M, Elliot GS, Sarmiento U, Senaldi G;
 PI WPI; 1998-437475/37.
 XX N-PSDB; AAV47512.
 DR Newly isolated nucleic acid encoding human or murine neurotrophic
 XX factor NNT-1, useful for treatment of neurological and
 PT immunological diseases or inflammation, also as vaccine adjuvant
 PT Claim 13; Fig 5; 120pp; English.
 XX This is the amino acid sequence of a murine neurotrophic factor,
 CC designated NNT-1, that is a growth factor for neurons and for B or
 CC T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512).
 CC Human NNT-1 (see AAV29715) is also provided. Vectors and host cells
 CC for use in the production of human murine recombinant NNT-1
 CC polypeptides. These are used to treat: (i) neurological or
 CC immunological diseases, specifically Alzheimer's, Parkinson's
 CC or Huntington's diseases, amyotrophic lateral sclerosis,
 CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
 CC degeneration of the neural retina, or conditions characterised by T
 CC or B cell defects, e.g. common variable immunodeficiency (CVID),
 CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
 CC antibody production following vaccination, and, since it inhibits
 CC tumour necrosis factor production, it may also be useful for
 CC treating sepsis. In addition, cells that have been engineered to
 CC express NNT-1 can be implanted, or nucleic acids are delivered in
 CC gene therapy vectors.
 XX SQ Sequence 225 AA;

Query Match 100.0%; Score 1224; DB 19; Length 225;
 Best Local Similarity 100.0%; Pred. No. 2.4e-121;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGSPSIQKTYDLYLEHQLRSLAGT 60
 Db 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGSPSIQKTYDLYLEHQLRSLAGT 60
 QY 61 YLYLGPPEPPEDFPNPRLGAEITLPRATVNLVWRSINDRRLTONYEAYSHLLCYLRGL 120
 Db 61 YLYLGPPEPPEDFPNPRLGAEITLPRATVNLVWRSINDRRLTONYEAYSHLLCYLRGL 120
 QY 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180
 Db 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180
 QY 181 KMDDFWLLKELQTLWRSKDFNRLKKMQPPAASVTLLHLEAHGF 225
 Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMQPPAASVTLLHLEAHGF 225

RESULT 2
 AAW56142
 ID AAW56142 standard; Protein; 225 AA.
 XX AC AAW56142;
 XX 13-JUL-1998 (first entry)
 DT Amino acid sequence of murine neurotrophic factor NNT-1.
 DE Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
 KW treatment; neurological disease; degeneration; Parkinson's disease;
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
 XX Mus sp. *

XX Key Location/Qualifiers
 FH Peptide 1..27
 FT /note= "signal peptide"
 FT Protein 28..225
 FT /note= "mature peptide"
 XX US5741772-A.
 XX 21-APR-1998.
 XX 03-FEB-1997; 97US-0792019.
 XX 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 XX Chang M;
 XX WPI; 1998-260526/23.
 XX N-PSDB; AAV22654.
 XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
 PT useful for stimulating growth of motor and sympathetic neurons
 XX Claim 2; Fig 5; 41pp; English.
 XX The present sequence represents a murine neurotrophic factor, designated
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic
 CC neurons. The NNT-1 protein is useful in the treatment of neurological
 CC diseases characterised by the degeneration and death of particular
 CC classes of neurons. These diseases specifically include Parkinson's
 CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
 CC stroke and various degenerative disorders affecting vision.
 XX SQ Sequence 225 AA;

Query Match 100.0%; Score 1224; DB 19; Length 225;
 Best Local Similarity 100.0%; Pred. No. 2.4e-121;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGSPSIQKTYDLYLEHQLRSLAGT 60
 Db 1 MDLRAGDSWGMACLTVMHLPAVPALNRTGDPGSPSIQKTYDLYLEHQLRSLAGT 60
 QY 61 YLYLGPPEPPEDFPNPRLGAEITLPRATVNLVWRSINDRRLTONYEAYSHLLCYLRGL 120
 Db 61 YLYLGPPEPPEDFPNPRLGAEITLPRATVNLVWRSINDRRLTONYEAYSHLLCYLRGL 120
 QY 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180
 Db 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180
 QY 181 KMDDFWLLKELQTLWRSKDFNRLKKMQPPAASVTLLHLEAHGF 225
 Db 181 KMDDFWLLKELQTLWRSKDFNRLKKMQPPAASVTLLHLEAHGF 225

RESULT 3
 AAY87814
 ID AAY87814 standard; Protein; 225 AA.
 XX AC AAY87814;
 XX 24-AUG-2000 (first entry)
 DT Murine NNT-1 protein.
 DE NNT-1; neurotrophic factor; neurotrophic; neuroprotective; treatment;
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;

KW retinopathy; immune disorder; hematopoietic disorder.
 XX Mus sp.
 OS US6054294-A.
 PN 25-APR-2000.
 XX 12-DEC-1997; 97US-0988919.
 PD 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 PA Chang M;
 PI WPI; 2000-338492/29.
 DR N-PSDB; AAA39483.
 XX New nucleic acids encoding neurotrophic factors useful for stimulating
 PT growth of motor or sympathetic neurons for treating neuron cell damage
 PT
 XX
 PS Claim 2b; Fig 5; 42pp; English.
 XX This invention describes a novel nucleic acid molecule (I) encoding a
 CC novel neurotrophic factor (NNT-1) (II) which has neurotrophic,
 CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
 CC ophthalmological activity. (I) is useful for producing NNT-1
 CC polypeptides which are useful for treating patients in whom various
 CC cells of the central, autonomic, or peripheral nervous system have
 CC degenerated and/or have been damaged by congenital disease, trauma,
 CC mechanical damage, surgery, stroke, ischemia, infection, metabolic
 CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
 CC proteins are used to treat diseases like Alzheimer's, Parkinson's,
 CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
 CC disease, peripheral neuropathy induced by diabetes or other metabolic
 CC disorders, and/or dystrophies or degeneration of the neural retina such
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
 CC night blindness, progressive cone-rod degeneration, immune disorders and
 CC hematopoietic disorders. (I) is effective in treating neurological
 CC conditions and promotes neuron regeneration. Neural functions are
 CC effectively restored in patients suffering from various neurological
 CC disorders. This sequence represents the murine NNT-1 protein described in
 CC the method of the invention.
 XX
 SQ Sequence 225 AA;
 Query Match 100.0%; Score 1224; DB 21; Length 225;
 Best Local Similarity 100.0%; Pred. No. 2.4e-121;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGLACLTCLVTLWHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLSLAGT 60
 DB 1 MDLRAGDSWGLACLTCLVTLWHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLSLAGT 60
 QY 61 YLYNLGPPFNEPDPNPRLGASTLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
 DB 61 YLYNLGPPFNEPDPNPRLGASTLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
 QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPGTPEWAPGPAHSDFLQ 180
 DB 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMATLGYPLPQPLPGTPEWAPGPAHSDFLQ 180
 QY 181 KMDDFWLLKELQTLWRSKADFNRLKKQPPAASVTILHLEAHGF 225
 DB 181 KMDDFWLLKELQTLWRSKADFNRLKKQPPAASVTILHLEAHGF 225
 RESULT 4
 AAU78177
 ID AAU78177 standard; Protein; 225 AA.
 XX

AAU78177;
 05-JUN-2002 (first entry)
 Mouse novel neurotrophic factor NNT1.
 Mouse; NNT1; neurotrophic factor; IGE-related disease;
 Type I allergic disease; allergic rhinitis; eczema; dermatitis;
 pollinosis; asthma; immune disease; cancer; arteriosclerosis;
 vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
 inflammatory arthritis; osteoarthritis; inflammatory joint disease;
 autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
 inflammatory bowel disease; transplant rejection; reproductive disorder;
 graft versus host disease; infertility; miscarriage; preterm labour.
 Mus sp.
 WO200215977-A2.
 28-FEB-2002.
 17-AUG-2001; 2001WO-US25906.
 18-AUG-2000; 2000US-226436P.
 16-AUG-2001; 2001US-0931704.
 (AMGE-) AMGEN INC.
 Senaldi G;
 WPI; 2002-280967/32.
 N-PSDB; ABK11649.
 Treating Immunoglobulin E-related disease, modulating IGE levels in a
 patient, preventing IGE-related disease and treating allergic diseases,
 involves administering NNT-1 inhibitor to a patient
 Claim 2; Fig 5; 63pp; English.
 The invention relates to treating Immunoglobulin E (IGE)-related disease,
 modulating IGE levels in a patient, preventing an IGE-related disease,
 and treating allergic diseases, comprising administering a
 therapeutically effective amount of novel neurotrophic factor (NNT)-1
 inhibitor to a patient. Also included are a method of diagnosing an
 IGE-related disease or susceptibility to an IGE-related disease, by
 determining the presence or amount of expression of an NNT1 polypeptide
 encoded by a NNT1 nucleotide sequence, its fragment or naturally
 occurring variant, and diagnosing an IGE-related disease or
 susceptibility of an IGE-related disease based on the presence or amount
 of expression of the polypeptide and a pharmaceutical composition for use
 in treating IGE-related disease, comprising the NNT1 inhibitor.
 The NNT1 inhibitor is useful for preventing and treating IGE-related
 disease, modulating IGE levels, and treating allergic diseases e.g.
 Type I allergic disease, allergic rhinitis, eczema, dermatitis,
 pollinosis, asthma, immune diseases and disorders, diseases involving
 abnormal cell proliferation including cancer, arteriosclerosis and
 vascular restenosis, diseases and conditions relating to dysfunction of
 immune system including rheumatoid arthritis, psoriatic arthritis,
 inflammatory arthritis, osteoarthritis, inflammatory joint disease,
 autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
 bowel disease, transplant rejection, and graft versus host disease, and
 reproductive diseases and disorders including infertility, miscarriage,
 preterm labour and delivery, and endometriosis. The present sequence
 represents Mouse NNT1.
 Sequence 225 AA;
 Query Match 100.0%; Score 1224; DB 23; Length 225;
 Best Local Similarity 100.0%; Pred. No. 2.4e-121;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGLACLTCLVTLWHLPAVPALNRTGDPGPGPSIQKTYDLYLEHQLSLAGT 60
 |||

Db 1 MDLRAGDSWGMGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT 60
 QY 61 YLYNLGPPPEPDPNPRGLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120
 Db 61 YLYNLGPPPEPDPNPRGLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120
 QY 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTPEPAWPGPAHSDFLQ 180
 Db 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTPEPAWPGPAHSDFLQ 180
 QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAASVTLHLEAHGF 225
 Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAASVTLHLEAHGF 225

RESULT 5

AAW29715
 ID AAW29715 standard; Protein; 225 AA.

AC AAW29715;

DT 09-NOV-1998 (first entry)

DE Human neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
 KW peripheral neuropathy; dystrophy; neural retina degeneration;
 KW common variable immunodeficiency; CVID; selective IGA deficiency;
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
 KW therapy.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..27

FT Protein /label= Sig_peptide

FT /label= Mat_protein

XX WO9833922-A1.

XX 06-AUG-1998.

PD 02-FEB-1998; 98WO-US02363.

PF 30-JAN-1998; 98US-0016534.

PR 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M, Elliot GS, Sarmiento U, Senaldi G;

XX WPI; 1998-437475/37.

XX N-PSDB; AAV47510-11.

XX Newly isolated nucleic acid encoding human or murine neurotrophic

XX factor NNT-1 - useful for treatment of neurological and

XX immunological diseases or inflammation, also as vaccine adjuvant

XX Claim 12; Fig 3; 120pp; English.

XX This is the amino acid sequence of a novel neurotrophic factor,

XX designated NNT-1, that is a growth factor for neurons and for B or

XX T cells. It was deduced from isolated cDNA (see AAV47510) and

XX genomic DNA (see AAV47511) clones. Vectors containing the cDNA or

XX genomic DNA and host cells are provided for use in the production

XX of NNT-1 polypeptides. These are used to treat: (i) neurological

XX or immunological diseases, specifically Alzheimer's, Parkinson's

XX or Huntington's diseases, amyotrophic lateral sclerosis,

XX Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and

XX degeneration of the neural retina, or conditions characterised by T

CC or B cell defects, e.g. common variable immunodeficiency (CVID),
 CC selective IGA deficiency, hypogammaglobulinaemia and X-linked
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
 CC inflammation. NNT-1 is also able to boost immunoreactivity and
 CC antibody production following vaccination, and, since it inhibits
 CC tumour necrosis factor production, it may also be useful for
 CC treating sepsis. In addition, cells that have been engineered to
 CC express NNT-1 can be implanted, or nucleic acids are delivered in
 CC gene therapy vectors.

XX Sequence 225 AA;

Query Match 97.5%; Score 1193; DB 19; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDLRAGDSWGMGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT 60

Db 1 MDLRAGDSWGMGLACLTIVLWHLPAVPALNRTGDPGPGPSIQKTYDLTRYLHQLRSLAGT 60

QY 61 YLYNLGPPPEPDPNPRGLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120

Db 61 YLYNLGPPPEPDPNPRGLGAEITLPRATVNLVWRSNDRLRLTONYAYSHLLCYLRGL 120

QY 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTPEPAWPGPAHSDFLQ 180

Db 121 NRQAATAELRRSLAHFCTSLQGLLGSAGVMTLGYPLPQPLPGTPEPAWPGPAHSDFLQ 180

QY 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAASVTLHLEAHGF 225

Db 181 KMDDFWLLKELQTLWLRSAKDFNRLKKMQPPAAASVTLHLEAHGF 225

RESULT 6

AAW56141

ID AAW56141 standard; Protein; 225 AA.

XX AAW56141;

DT 13-JUL-1998 (first entry)

XX Amino acid sequence of human neurotrophic factor NNT-1.

XX Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
 KW treatment; neurological disease; degeneration; Parkinson's disease;
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..27

FT Protein /note= "signal peptide"

FT /note= "mature protein"

XX US5741772-A.

XX 21-APR-1998.

XX 03-FEB-1997; 97US-0792019.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M;

XX WPI; 1998-260526/23.

XX N-PSDB; AAV22652.

XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -

XX useful for stimulating growth of motor and sympathetic neurons

PS Claim 1; Fig 3; 4lpp; English.

XX The present sequence represents a human neurotrophic factor, designated

CC NNT-1, which is capable of stimulating growth of motor or sympathetic

CC neurons. The NNT-1 protein is useful in the treatment of neurological

CC diseases characterised by the degeneration and death of particular

CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,

CC stroke and various degenerative disorders affecting vision.

XX Sequence 225 AA;

SQ Sequence 225 AA;

Query Match 97.5%; Score 1193; DB 19; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMCLACTVLMHLPAPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Db 1 MDLRAGDSWGMCLACTVLMHLPAPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Qy 61 YLYNLGPPFPNPDNPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Db 61 YLYNLGPPFPNPDNPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Qy 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Db 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

RESULT 7

AAW94466

ID AAW94466 standard; Protein; 225 AA.

XX AAW94466;

AC AAW94466;

XX 22-APR-1999 (first entry)

DT Human cardiotrophin-like cytokine protein.

DE Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;

XX CXC; IL-6; diagnosis; detection; immune system-related disorder;

KW cancer; cardiac disorder; heart failure; hypertension; cancer;

KW autoimmune disorder; infection.

XX Homo sapiens.

OS Homo sapiens.

PN WO9900415-A1.

XX 07-JAN-1999.

PD 29-JUN-1998; 98WO-US13129.

PF 30-JUN-1997; 97US-0051311.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Shi Y;

PI WPI; 1999-095678/08.

XX N-PSDB; AAX16161.

DR New isolated cardiotrophin-like cytokine nucleic acid - used to

XX develop products for treating cardiac and immune system disorders,

PT e.g. heart failure, hypertension, cancers, autoimmune disorders and

PT infections

XX Claim 1; Fig 1; 103pp; English.

PS The present invention relates to a novel cardiotrophin-like cytokine

XX (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine

CC family. The present sequence represents the human CLC protein. The

CC present invention also describes screening methods for identifying

CC agonists and antagonists of CLC activity, as well as methods for

CC detecting cardiac and immune system-related disorders and

CC therapeutic methods for treating cardiac and immune system-related

CC disorders, e.g. heart failure, hypertension, cancers, autoimmune

CC disorders and infections.

XX SQ Sequence 225 AA;

Query Match 97.5%; Score 1193; DB 20; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;

Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGMCLACTVLMHLPAPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Db 1 MDLRAGDSWGMCLACTVLMHLPAPVAPALNRTGDPGPGPSIQKTYDLTRYLHQLSLAGT 60

Qy 61 YLYNLGPPFPNPDNPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Db 61 YLYNLGPPFPNPDNPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120

Qy 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Db 121 NRQAATAEILRRSLAHFCTSLQGLGSIAGVMATLGYPLPQPLPGTEPAWAPGPAHSDFLQ 180

Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLEAHGF 225

RESULT 8

AAW87813

ID AAW87813 standard; Protein; 225 AA.

XX AAW87813;

AC AAW87813;

XX 24-AUG-2000 (first entry)

DT Human NNT-1 protein.

DE Human NNT-1 protein.

XX NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;

KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;

KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;

KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;

KW Huntington's disease; peripheral neuropathy; neural retina degeneration;

KW retinopathy; immune disorder; hematopoietic disorder.

XX Homo sapiens.

OS Homo sapiens.

PN US6054294-A.

XX 25-APR-2000.

PD 12-DEC-1997; 97US-0988819.

PF 03-FEB-1997; 97US-0792019.

XX

XX (AMGE-) AMGEN INC.
 XX Chang M;
 XX WPI: 2000-338492/29.
 DR N-PSDB; AAA39481.
 XX
 PT New nucleic acids encoding neurotrophic factors useful for stimulating
 PT growth of motor or sympathetic neurons for treating neuron cell damage
 XX
 PS Claim 1c; Fig 3; 42pp; English.
 XX
 CC This invention describes a novel nucleic acid molecule (I) encoding a
 CC novel neurotrophic factor (NNT-1) (II) which has neurotrophic, neuroprotective,
 CC anticonvulsant, antiparkinsonian, antidiabetic and ophthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic disorders, and/or dystrophies or degeneration of the neural retina such as retinitis pigmentosa, drug-induced retinopathies, stationary forms of night blindness, progressive cone-rod degeneration, immune disorders and hematopoietic disorders. (II) is effective in treating neurological conditions and promotes neuron regeneration. Neural functions are effectively restored in patients suffering from various neurological disorders. This sequence represents the human NNT-1 protein described in the method of the invention.
 XX
 SQ Sequence 225 AA;
 Query Match 97.5%; Score 1193; DB 21; Length 225;
 Best Local Similarity 96.9%; Pred. No. 4.7e-118;
 Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGLACLTVMHLPAVPAALNRTGDPGPGSIQKTYDLYLHQLRSLAGT 60
 Db 1 MDLRAGDSWGLACLTVMHLPAVPAALNRTGDPGPGSIQKTYDLYLHQLRSLAGT 60
 QY 61 YLNYLGPPFPNEPDPNPPRLGAETLPRAVPAALNRTGDPGPGSIQKTYDLYLHQLRSL 120
 Db 61 YLNYLGPPFPNEPDPNPPRLGAETLPRAVPAALNRTGDPGPGSIQKTYDLYLHQLRSL 120
 QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAATGYPLPQPLGTEPAPAGPSHDFLQ 180
 Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAATGYPLPQPLGTEPAPAGPSHDFLQ 180
 QY 181 KMDDFWLLKELQTLWRSKDNRLKKMKOPPAASVTLHLEAHGF 225
 Db 181 KMDDFWLLKELQTLWRSKDNRLKKMKOPPAASVTLHLEAHGF 225
 RESULT 9
 AAG63543
 ID AAG63543 standard; Protein; 225 AA.
 XX
 AC AAG63543;
 XX
 DT 15-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of a human NNT-1 protein.
 XX
 KW NNT-1; CLF-1; SCNTFRalpha; nervous system; neuron; nervous system;
 KW neuro-muscular function; tumour; immune system; haematopoietic system;
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;

KW blastocyst implantation; thrombosis; retinal disease;
 KW retinal pigmentosis.
 XX Homo sapiens.
 XX WO200155172-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 26-JAN-2001; 2001WO-FR00253.
 PF
 XX 27-JAN-2000; 2000FR-0001035.
 PR 12-OCT-2000; 2000FR-0013089.
 XX
 XX (FABR) FABRE MEDICAMENT SA PIERRE.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;
 WPI: 2001-488773/53.
 N-PSDB; AAH74484.
 XX
 CC A complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha protein useful to treat neurodegenerative disease including Parkinson's and Huntington's, obesity and cancer -
 PT
 PT Claim 2; Page 58; 67pp; French.
 XX
 CC The present sequence represents a human NNT-1 protein. The specification describes a complex comprising a NNT-1 protein and a CLF-1 and/or SCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate activity of the SCNTFRalpha/gp130/LiPRbeta receptor complex, or to induce phosphorylation of the tyrosine of gp130 and LiPRbeta, particularly where cells expressing the receptor complex are in the central or peripheral nervous system, in neurons implicated in neuro-muscular function or in skeletal muscle. The complex or antibodies are also used to decrease the survival, growth or proliferation of tumour cells or to facilitate the proliferation and/or inhibit differentiation of cells stocks. The complex is also used to modulate activity of the gp130/LiPRbeta receptor or cells expressing haematopoietic, nervous or reproductive system, the liver or skeletal muscle. Molecules of the invention may be used to prevent or treat neurodegenerative diseases including amyotrophic lateral sclerosis, Parkinson's and Huntington's disease, to repair or regenerate nervous or muscular tissue or to maintain muscular mass in paralysis patients. They may also be used to treat cancer, obesity and associated diseases, and to improve fertility, particularly to avoid endometriosis and/or assist blastocyst implantation, thrombosis, or retinal disease, particular retinal pigmentosis.
 XX
 SQ Sequence 225 AA;
 Query Match 97.5%; Score 1193; DB 22; Length 225;
 Best Local Similarity 96.9%; Pred. No. 4.7e-118;
 Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MDLRAGDSWGLACLTVMHLPAVPAALNRTGDPGPGSIQKTYDLYLHQLRSLAGT 60
 Db 1 MDLRAGDSWGLACLTVMHLPAVPAALNRTGDPGPGSIQKTYDLYLHQLRSLAGT 60
 QY 61 YLNYLGPPFPNEPDPNPPRLGAETLPRAVPAALNRTGDPGPGSIQKTYDLYLHQLRSL 120
 Db 61 YLNYLGPPFPNEPDPNPPRLGAETLPRAVPAALNRTGDPGPGSIQKTYDLYLHQLRSL 120
 QY 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAATGYPLPQPLGTEPAPAGPSHDFLQ 180
 Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMAATGYPLPQPLGTEPAPAGPSHDFLQ 180
 QY 181 KMDDFWLLKELQTLWRSKDNRLKKMKOPPAASVTLHLEAHGF 225
 Db 181 KMDDFWLLKELQTLWRSKDNRLKKMKOPPAASVTLHLEAHGF 225

RESULT 10
AAU78176
ID AAU78176 standard; Protein; 225 AA.
XX AC
XX AAU78176;
XX
XX 05-JUN-2002 (first entry)
XX DE Human novel neurotrophic factor NNT1.
XX
XX Human; NNT1; neurotrophic factor; IGE-related disease;
KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;
KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
KW inflammatory bowel disease; transplant rejection; reproductive disorder;
KW graft versus host disease; infertility; miscarriage; preterm labour.
XX
OS Homo sapiens.
XX
XX W0200215977-A2.
PN
XX
XX 28-FEB-2002.
PD
XX
XX 17-AUG-2001; 2001WO-US25906.
PF
XX
XX 18-AUG-2000; 2000US-226436P.
PR
XX 16-AUG-2001; 2001US-0931704.
PR
XX
XX (AMGE-) AMGEN INC.
PA
XX
XX Senaldi G;
PI
XX WPI; 2002-280967/32.
DR
XX N-PSDB; ABK11647.
DR
XX
XX Treating Immunoglobulin E-related disease, modulating IGE levels in a
PT patient, preventing IGE-related disease and treating allergic diseases,
PT involves administering NNT-1 inhibitor to a patient -
XX
XX Claim 2; Fig 3; 63pp; English.
PS
XX The invention relates to treating Immunoglobulin E (IGE)-related disease,
CC modulating IGE levels in a patient, preventing an IGE-related disease,
CC and treating allergic diseases, comprising administering a
CC therapeutically effective amount of novel neurotrophic factor (NNT)-1
CC inhibitor to a patient. Also included are a method of diagnosing an
CC IGE-related disease or susceptibility to an IGE-related disease, by
CC determining the presence or amount of expression of an NNT1 polypeptide
CC encoded by a NNT1 nucleotide sequence, its fragment or naturally
CC occurring variant, and diagnosing an IGE-related disease or
CC susceptibility of an IGE-related disease based on the presence or amount
CC of expression of the polypeptide and a pharmaceutical composition for use
CC in treating IGE-related disease, comprising the NNT1 inhibitor.
CC The NNT1 inhibitor is useful for preventing and treating IGE-related
CC disease, modulating IGE levels, and treating allergic diseases e.g.
CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,
CC pollinosis, asthma, immune diseases and disorders, diseases involving
CC abnormal cell proliferation including cancer, arteriosclerosis and
CC vascular restenosis, diseases and conditions relating to dysfunction of
CC immune system including rheumatoid arthritis, psoriatic arthritis,
CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,
CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
CC bowel disease, transplant rejection, and graft versus host disease, and
CC reproductive diseases and disorders including infertility, miscarriage,
CC preterm labour and delivery, and endometriosis. The present sequence
CC represents human NNT1.
XX
SQ Sequence 225 AA;
Query Match 97.5%; Score 1193; DB 23; Length 225;

Best Local Similarity 96.9%; Pred. No. 4.7e-118;
Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 MDLRAGDSGMLACICTVLWHLPAVPALNRTGDPGPSIOKTYDLTRYLEHQLRSLAGT 60
DB 1 MDLRAGDSGMLACICTVLWHLPAVPALNRTGDPGPSIOKTYDLTRYLEHQLRSLAGT 60
QY 61 YLNYLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
DB 61 YLNYLGPPFNEPDNPPRLGAETLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
QY 121 NRQAATAELRRSLAHFCTSLQGLLSGAGVATLGYPLPQLPCTEPAWAGPAHSDFLQ 180
DB 121 NRQAATAELRRSLAHFCTSLQGLLSGAGVATLGYPLPQLPCTEPAWAGPAHSDFLQ 180
QY 181 KMDDFWLLKELQTLWRSKDFNLRKKMOPPAASVTLHLEAHGF 225
DB 181 KMDDFWLLKELQTLWRSKDFNLRKKMOPPAASVTLHLEAHGF 225
RESULT 11
AAU25831
ID AAU25831 standard; Protein; 253 AA.
XX AC
XX AAU25831;
XX
XX 16-OCT-2001 (first entry)
DE Human protein sequence SEQ ID NO:1346.
XX
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antidiabetic; osteopathic; eczema;
KW dermatological; anti-allergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW immunophylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX
XX Homo sapiens.
OS
XX W0200153455-A2.
PN
XX 26-JUL-2001.
PD
XX 22-DEC-2000; 2000WO-US35017.
PF
XX 23-DEC-1999; 99US-0471275.
PR
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-457603/49.
XX N-PSDB; AAH99772.
DR
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 20; Page 278; 1217pp; English.
PS
XX AAH99166 to AAH99904 encode the human proteins given in AAU25225 to
CC AAU25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;

antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 central nervous system; viricide; anti-HIV; fungicide; antimutagen;
 cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 encoding them can be used in gene therapy, antisense therapy and vaccine
 production. The proteins and polynucleotides are useful for screening for
 agonists or antagonists of a protein and for the treatment and diagnosis
 of disorders associated with the activity of a protein e.g. inflammation,
 rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 infections, autoimmunity, genetic diseases, haematopoietic disorders,
 anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 osteoporosis, severe combined immunodeficiency, eczema, allergic
 rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 Alzheimer's disease, Parkinson's disease, neurodegenerative and
 neurological disorders.

Sequence 253 AA;
 Query Match 97.5%; Score 1193; DB 22; Length 253;
 Best Local Similarity 96.9%; Pred. No. 5.5e-118;
 Matches 218; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTVMWHPALNPALNRTGDPGPGSIQKTYDTRYLHQLRSLAGT 60
 Db 29 MDLRAGDSWGLACLTVMWHPALNPALNRTGDPGPGSIQKTYDTRYLHQLRSLAGT 88
 Qy 61 YLYNLGPPNEPDPNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
 Db 89 YLYNLGPPNEPDPNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 148
 Qy 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTPEPAWAPGPAHSDFLQ 180
 Db 149 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTPEPAWAPGPAHSDFLQ 208
 Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLGAHF 225
 Db 209 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLGAHF 253

RESULT 12
 AAE00828
 ID AAE00828 standard; Protein; 223 AA.

AAE00828;
 02-JUL-2001 (first entry)
 Human cardiotrophin-like cytokine (CLC) protein.
 Human; biologically active complex; haemopoietin receptor; NR6;
 cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;
 differentiation; cell survival; neurotrophic activity.
 Homo sapiens.
 Key Location/Qualifiers
 Peptide 1..27
 /label= Signal_peptide
 Protein 28..223
 /label= Human_mature_CLC_protein
 /note= "Cardiotrophin-like cytokine"

WO200127157-A1.
 19-APR-2001.
 06-OCT-2000; 2000WO-AU01216.
 08-OCT-1999; 99AU-0003327.
 12-MAY-2000; 2000AU-0007489.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 PA Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;
 PI Nakata Y, Hasegawa M;
 XX WPI; 2001-281978/29.
 DR N-PSDB; AAD04201.
 XX New biologically active complex comprising NR6 and
 PT cardiotrophin-like-cytokine, for facilitating proliferation,
 PT differentiation and/or survival of a cell -
 XX Claim 32; Page 114-115; 123pp; English.
 XX The present invention relates to a biologically active complex comprising
 CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).
 CC The complex is useful in the manufacture of a medicament for the
 CC treatment and/or prophylaxis of a subject, as it is involved in
 CC facilitating proliferation, differentiation and/or survival of a cell.
 CC The complex or its components have neurotrophic activity. The present
 CC sequence is human cardiotrophin-like cytokine (CLC) protein.

Sequence 223 AA;
 Query Match 96.5%; Score 1181; DB 22; Length 223;
 Best Local Similarity 96.9%; Pred. No. 8.7e-117;
 Matches 216; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDLRAGDSWGLACLTVMWHPALNPALNRTGDPGPGSIQKTYDTRYLHQLRSLAGT 60
 Db 1 MDLRAGDSWGLACLTVMWHPALNPALNRTGDPGPGSIQKTYDTRYLHQLRSLAGT 60
 Qy 61 YLYNLGPPNEPDPNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
 Db 61 YLYNLGPPNEPDPNPRLGAEITLPATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL 120
 Qy 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTPEPAWAPGPAHSDFLQ 180
 Db 121 NRQAATAELRRSLAHFCTSLQGLLSIAGVMTLGYPLPQLPGTPEPAWAPGPAHSDFLQ 180
 Qy 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLGAH 223
 Db 181 KMDDFWLLKELOTWLRSAKDFNRLKKMKQPPAAAVTLHLGAH 223

RESULT 13
 ABB11896
 ID ABB11896 standard; peptide; 321 AA.

ABB11896;
 11-JAN-2002 (first entry)
 Human cardiotrophin-like cytokine homologue, SEQ ID NO:2266.
 Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; athma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerary; antiulcer.

Homo sapiens.
 OS
 XX WO200157188-A2.
 PN

XX PD 09-AUG-2001.
 XX PF 05-FEB-2001; 2001WO-US03800.
 XX PR 03-FEB-2000; 2000US-0496914.
 XX PR 27-APR-2000; 2000US-0560875.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX DR WPI; 2001-457740/49.
 XX DR N-PSDB; ABA09140.
 XX PT Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 XX PS Claim 20; Page 273; 1963pp; English.
 XX CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.

XX Sequence 321 AA;

Query Match 95.7%; Score 1171; DB 22; Length 321;
 Best Local Similarity 96.0%; Pred. No. 1.6e-115;
 Matches 214; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 LRAGDSWGMACLCVTLVHLPVAPALNRTGDPGPGPSIQKTVLDLTRYLHQLRSLAGTYL 62

Db 99 LPTGDSWGMACLCVTLVHLPVAPALNRTGDPGPGPSIQKTVLDLTRYLHQLRSLAGTYL 158

QY 63 NYLGPFPNPDNPPRLGHTLPRATVNLVWRSINDRLRLTQNYEAYSHLLCYLRGLNR 122

Db 159 NYLGPFPNPDNPPRLGHTLPRATVNLVWRSINDRLRLTQNYEAYSHLLCYLRGLNR 218

QY 123

QY 219

QY 183

Db 279

RESULT 14

AAM79399

ID AAM79399 standard; Protein; 321 AA.

XX AC AAM79399;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3045.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX PI Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX DR WPI; 2001-476283/51.

XX DR N-PSDB; AAK52532.

XX PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -

XX PS Claim 20; Page 237; 6221pp; English.

XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

XX CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX SQ Sequence 321 AA;

Query Match 95.7%; Score 1171; DB 22; Length 321;

Best Local Similarity 96.0%; Pred. No. 1.6e-115;
Matches 214; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 LRAGDSWGLACLTCTVHLPAVPALNRTGDPGPGPSIOKTYDLTRYLEHQLRSLAGTYL 62
DB 99 LFTGDSWGLACLTCTVHLPAVPALNRTGDPGPGPSIOKTYDLTRYLEHQLRSLAGTYL 158
QY 63 NYLGPPFNEPDPNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNR 122
DB 159 NYLGPPFNEPDPNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNR 218
QY 123 QAATAELRSLAHFCTSLQGLLSIAGVWATLGYPLPQPLPGTEPAWAPGPAHSDFLOKM 182
DB 219 QAATAELRSLAHFCTSLQGLLSIAGVWATLGYPLPQPLPGTEPTWTPGPAHSDFLOKM 278
QY 183 DDFWLLKELQTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225
DB 279 DDFWLLKELQTLWRSKDFNRLKKMQPPAAAVTLHLEAHGF 321

RESULT 15

AAB19587

ID AAB19587 standard; Protein; 215 AA.

XX AAB19587;

AC AAB19587;
DT 22-JAN-2001 (first entry)
XX Mouse interleukin-B60 (IL-B60).

XX Interleukin-B60; IL-B60; mouse; cytokine; cytokine-like factor-1;

KW haematopoietic; inflammation; antiinflammatory; autoimmune disease;
KW therapy.
XX Mus musculus.

XX Key Location/Qualifiers

FT Peptide 1..17

FT /label= Signal_peptide

FT Protein 18..215

FT /label= Mature-protein

XX WO2000053631-A1.

XX 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US06182.

XX 11-MAR-1999; 99US-0267901.

XX (SCHE) SCHERING CORP.

XX Oppmann B, Timans JC, Kastelein RA, Bazan JF;

XX WPI; 2000-587426/55.

XX N-PSDB; AAA88547.

XX Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes,

XX polypeptides, and nucleic acids, useful in research, diagnosis and for

XX treating inflammatory and autoimmune disorders -

XX Claim 1; Page 17; 97pp; English.

XX The present sequence is that of mouse interleukin-B60 (IL-B60), a

XX novel, small soluble cytokine-like protein that exhibits structural

XX motifs characteristic of a member of the long-chain cytokines, and

XX which shows homology to granulocyte colony stimulating factor and

XX interleukin-6. IL-608 may have either stimulatory or inhibitory

XX effects on haematopoietic cells, including e.g. lymphoid cells,

XX such as T-cells, B-cells, natural killer cells, macrophages,

XX dendritic cells, haematopoietic progenitors, etc. Methods are

XX provided for modulating the physiology or development of a cell or

XX tissue culture cells by contacting the cell with an agonist or

CC antagonist of IL-B60 or an agonist of antagonist of a complex of
CC mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see
CC AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological
CC factor in motor neuron development and regeneration. IL-608, its
CC agonists and antagonists may be used to treat inflammatory or
CC autoimmune disorders and also for drug screening.

XX
SQ Sequence 215 AA;

Query Match 95.3%; Score 1167; DB 21; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5e-115;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 MLACLCVTLVHLPAVPALNRTGDPGPGPSIOKTYDLTRYLEHQLRSLAGTYLNYLGPPFN 70
DB 1 MLACLCVTLVHLPAVPALNRTGDPGPGPSIOKTYDLTRYLEHQLRSLAGTYLNYLGPPFN 60
QY 71 EPDFNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNRQAATAELR 130
DB 61 EPDFNPPRLGAETLPRATVNLVWRSNDRLRLTQNYEAYSHLLCYLRGLNRQAATAELR 120
QY 131 RSLAHFCTSLQGLLSIAGVWATLGYPLPQPLPGTEPAWAPGPAHSDFLOKMDDFWLLKE 190
DB 121 RSLAHFCTSLQGLLSIAGVWATLGYPLPQPLPGTEPAWAPGPAHSDFLOKMDDFWLLKE 180
QY 191 LQTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 225
DB 181 LQTLWRSKDFNRLKKMQPPAAASVTLHLEAHGF 215

Search completed: January 27, 2003, 15:41:34
Job time : 33 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 27, 2003, 15:42:42 ; Search time 9 Seconds
(without alignments)
504.464 Million cell updates/sec

Title: US-09-931-704-5

Perfect score: 1224

Sequence: 1 MDLRAGDSWGLACLTCLVLM.....KKKQPPAASVTLLHLEAHGF 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pdb.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pdb.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pdb.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pdb.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pdb.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pdb.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pdb.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pdb.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pdb.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pdb.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pdb.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pdb.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1224	100.0	225	10	US-09-931-704-5
2	1193	97.5	225	10	US-09-931-704-2
3	852	69.6	164	10	US-09-864-761-40014
4	110.5	9.0	203	10	US-09-896-856-3
5	92.5	7.6	243	9	US-10-000-776-6
6	92.5	7.6	243	9	US-09-791-497-8
7	92.5	7.6	243	10	US-09-810-052-5
8	91.5	7.5	201	10	US-09-901-540-3
9	91.5	7.5	201	10	US-09-896-856-8
10	91.5	7.5	201	10	US-09-901-257-3
11	90	7.4	195	10	US-09-770-361-5
12	87	7.1	242	9	US-10-000-776-2
13	87	7.1	242	9	US-09-791-497-2
14	86	7.0	232	10	US-09-810-052-2
15	81.5	6.7	348	10	US-09-730-617-2
16	80.5	6.6	379	9	US-09-860-846-16
17	80.5	6.6	379	10	US-09-861-289-16
18	80.5	6.6	3782	9	US-09-860-846-4
19	80.5	6.6	3782	10	US-09-861-289-4

20	79	6.5	1620	10	US-09-827-949-2	Sequence 2, Appli
21	78.5	6.4	1399	9	US-09-388-221-4	Sequence 4, Appli
22	78.5	6.4	1424	9	US-09-388-221-12	Sequence 12, Appli
23	78.5	6.4	1429	10	US-09-996-617-2	Sequence 2, Appli
24	78.5	6.4	1429	10	US-09-931-071-2	Sequence 2, Appli
25	78.5	6.4	1443	9	US-09-388-221-6	Sequence 6, Appli
26	78.5	6.4	1454	9	US-09-388-221-10	Sequence 10, Appli
27	78.5	6.4	1473	9	US-09-388-221-2	Sequence 2, Appli
28	78	6.4	427	8	US-08-808-031A-40	Sequence 40, Appli
29	78	6.4	625	10	US-09-771-161A-242	Sequence 242, App
30	78	6.4	625	10	US-09-771-161A-243	Sequence 243, App
31	78	6.4	917	10	US-09-815-242-5603	Sequence 5603, Ap
32	78	6.4	920	10	US-09-815-242-12181	Sequence 12181, A
33	78	6.4	920	10	US-09-815-242-12995	Sequence 12995, A
34	78	6.4	920	10	US-09-815-242-13148	Sequence 13148, A
35	77.5	6.3	218	10	US-09-893-737-28	Sequence 28, Appl
36	77.5	6.3	451	10	US-09-764-864-1344	Sequence 1344, Ap
37	77	6.3	473	10	US-09-871-874-19	Sequence 19, Appl
38	76	6.2	418	9	US-09-946-807-3	Sequence 3, Appli
39	76	6.2	418	10	US-09-795-668-3	Sequence 3, Appli
40	76	6.2	418	10	US-09-795-668-3	Sequence 3, Appli
41	75.5	6.2	237	10	US-09-925-301-867	Sequence 867, App
42	75.5	6.2	334	10	US-09-953-342-24	Sequence 24, Appl
43	74.5	6.1	1009	8	US-08-987-689A-2	Sequence 2, Appli
44	74	6.0	416	9	US-10-124-429-2	Sequence 2, Appli
45	73.5	6.0	856	9	US-09-364-847-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1

US-09-931-704-5
; Sequence 5, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating IGE-Related Disease Using
; TITLE OF INVENTION: Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Murine
US-09-931-704-5

Query Match	100.0%;	Score	1224;	DB	10;	Length	225;
Best Local Similarity	100.0%;	Pred. NO.	7.6e-112;				
Matches	225;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MDLRAGDSWGLACLTCLVHLPAVPALNRGTGDPGSPSIQKTYDLYLRYLHQLRSLAGT	60				
Db	1	MDLRAGDSWGLACLTCLVHLPAVPALNRGTGDPGSPSIQKTYDLYLRYLHQLRSLAGT	60				
Qy	61	YLYNLGPPFPNPDNPPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL	120				
Db	61	YLYNLGPPFPNPDNPPRLGAEITLPRATVNLVWRSNDRLRLTONYEAYSHLLCYLRGL	120				
Qy	121	NRQATAEIIRSLAHFCTSLQGLGSIAGVNTATGYPPLPQPLPGTEPAWAPGAHSDFLQ	180				
Db	121	NRQATAEIIRSLAHFCTSLQGLGSIAGVNTATGYPPLPQPLPGTEPAWAPGAHSDFLQ	180				
Qy	181	KMDDFWLLKELQTLWRSKDFNRLKKKQPPAASVTLLHLEAHGF	225				
Db	181	KMDDFWLLKELQTLWRSKDFNRLKKKQPPAASVTLLHLEAHGF	225				

;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006663	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006668	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006663	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006662	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006661	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	PCT/US01/006670	
;	PRIOR FILING DATE:	2001-01-30	
;	PRIOR APPLICATION NUMBER:	US 60/234,687	
;	PRIOR FILING DATE:	2000-09-21	
;	PRIOR APPLICATION NUMBER:	US 09/608,408	
;	PRIOR FILING DATE:	2000-06-30	
;	PRIOR APPLICATION NUMBER:	US 09/774,203	
;	PRIOR FILING DATE:	2001-01-29	
;	NUMBER OF SEQ ID NOS:	49117	
;	SOFTWARE:	Arnomax Sequence Listing Engine	
;	SEQ ID NO	40014	
;	LENGTH:	164	
;	TYPE:	PRT	
;	ORGANISM:	Homo sapiens	
;	FEATURE:		
;	OTHER INFORMATION:	MAP TO AC005849.1	
;	OTHER INFORMATION:	EXPRESSED IN BRAIN, S	
;	OTHER INFORMATION:	EXPRESSED IN BONE MAR	
;	OTHER INFORMATION:	EXPRESSED IN LUNG, S	
;	OTHER INFORMATION:	EXPRESSED IN ADULT L	
;	OTHER INFORMATION:	EXPRESSED IN PLACENT	
;	OTHER INFORMATION:	EXPRESSED IN HEART, S	
;	OTHER INFORMATION:	EXPRESSED IN HELA, S	
;	OTHER INFORMATION:	EXPRESSED IN FETAL L	
;	OTHER INFORMATION:	EST_HUMAN HIT: AI75292	
;	OTHER INFORMATION:	SWISSPROT HIT: Q63080	
;	US-09-864-761-40014		
	Query Match	69.6%;	Score 852
	Best Local Similarity	95.1%;	Pred. No.
	Matches	156;	Conservative
			4; Mismatch
QY	62	LNLYGPPNEPDPFNPRLGAETLP	LNLYGPPNEPDPFNPRLGAETLP
Db	1	LNLYGPPNEPDPFNPRLGAETLP	LNLYGPPNEPDPFNPRLGAETLP
QY	122	ROATAEURLRSLAHFCTSLQGL	ROATAEURLRSLAHFCTSLQGL
Db	61	ROATAEURLRSLAHFCTSLQGL	ROATAEURLRSLAHFCTSLQGL
QY	182	MDDFWLLKELQTLWRSKDNFNR	MDDFWLLKELQTLWRSKDNFNR
Db	121	MDDFWLLKELQTLWRSKDNFNR	MDDFWLLKELQTLWRSKDNFNR
	RESULT 4		
	US-09-896-856-3		
	Sequence 3, Application US/09896856		
	Patent No. US20020137189A1		
	GENERAL INFORMATION:		
	APPLICANT:	Baker, Joffre	
		Chien, Kenneth	
		King, Kathleen	
		Pennica, Diane	
		Wood, William	
	TITLE OF INVENTION:	Cardiac Hypertro	
	NUMBER OF SEQUENCES:	8	
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE:	Genentech, Inc.	
	STREET:	460 Point San Bruno Bl	
	CITY:	South San Francisco	
	STATE:	California	


```

; APPLICANT: Lawrence, David A.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret A.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the
; TITLE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P2533D2
; CURRENT APPLICATION NUMBER: US/09/901,540
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 09/648,258
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 09/033,114
; PRIOR FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: US 08/733,850
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: US 08/443,129
; PRIOR FILING DATE: 1995-05-17
; PRIOR APPLICATION NUMBER: US 08/286,304
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: US 08/233,609
; PRIOR FILING DATE: 1994-04-25
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Human
US-09-901-540-3

```

	Query Match	7.5%;	Score 91.5;	DB 10;	Length 201;
	Best Local Similarity	25.7%;	Pred. No. 0.1;		
	Matches 45;	Conservative 23;	Mismatches 92;	Indels 15;	Gaps
Qy	40	IKQTYDLTYLEHQLSLACTYLVNYLGPPNEPDEPPRLGAETLPATVNVLEWVRSLND	99		
Db	27	IRQTHSLAHLITKYAQLQEQYVQLQDPPGLFSFPPLFVAGL-SAPAPSHAGLFPVHE	85		
Qy	100	RURLTQNYEAYSHLLCYLRGLNRQAA-----TAEIERSLAHFCTSLQGLLSAGVNMATL	154		
Db	86	RURL--DAAALAPFLDDAVCRQRQELNPRAFLRLRUEDAARQARALGAIVALLAAL	141		
Qy	155	GY-----PLPQPLPGTEPAWAPGAHSDFTKQMDFWLLKELOTWLRSAKDFNRL	205		
Db	144	GAANRGPRAEPPAATAASA---ASATGVFPAKVLGLRVCGLYRWLSRSTEGDLGL	195		

1	MDLRAGD-SWGMLACLCTVL-----WHLPAVPALNRTGDPGPGPSIQK-----TYDL	46
QY		
1	MGQTADGLGWRLSLILLPLLVQAGVWGVPFRPPG-----RPQLSLQELRREFTVSLHL	53
Ddb		
47	TYLHEQLHSLAGTYLN-----YLGPPFNEPDNPPRLGNETLPRATVNLGVWRSLN	98
QY		
54	ARKLLSEVQGRHRAESHLPGVNLVLLP-----LG-EQDPDVSLTFQAWRRUS	101
Ddb		
99	DLRLTQNYEAYSHLCYLRGLNQRQAATAELRR-SLAHFCTSLQGLGSLAGVMTALGVP	157
QY		
102	DPERLCFISTTLQPHAPLGGGLCTGRWTNMERMQLWAMRLDLRLQRLHREQVLAAGFN	161
Ddb		
158	LPQP-----LPCTEPAWAGPAHSDFLOKMDDFWLLKELQTLWIRSAKDF	202
QY		
162	LPFEEREEEEERKGLLPGLGALSALQGPQVFPQLLSTYRLHLSLELVLSRAVEL	221
Ddb		
203	NRLKK	207
QY		
222	LLLSK	226
Ddb		

US-09-901-540-3
; Sequence 3, Application US/09901540
; Patent No. US2002010262A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: "Goddard, Audrey

US-09-896-856-8

US 09-036-030-8
; Sequence 8, Application US/09896856
; Patent No. US20020137189A1
; GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: Baker, Joffre
Chien, Kenneth

Chren, Kenneth
King, Kathleen
Pennica, Diane
Wood, William

MOORE, WILLIAM

TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd

STREET: 400 FINE SAN BLAS BLVD
CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb f

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatin (Genentech)

;;
CURRENT APPLICATION DATA:
APPLICATION NUMBER 10/00/000000

; FILING DATE: 29-JUN-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/733,850
; FILING DATE: 18-OCT-1996
; APPLICATION NUMBER: US 08/471,112
; FILING DATE: 06-JUN-1995
; APPLICATION NUMBER: 08/233,609
; FILING DATE: 25-APR-1994
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: 08/443129
; FILING DATE: 17-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P0894P1D2C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2066
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
;
; US-09-896-856-8

Query Match 7.5%; Score 91.5; DB 10; Length 201;
Best Local Similarity 25.7%; Pred. No. 0.1;
Matches 45; Conservative 23; Mismatches 92; Indels 15; Gaps 5;

QY 40 IQKTYDITRYLEHQLRSLAGTYLNYLGGPNEPDPRLGAETLPRTATVNVLEWRSIND 99
DB 27 IROTHSLAHLTKYAEQLQYVQLQDPPGLPSFSPRLPVAGL-SAPAPSHAGLPVHE 85
QY 100 RLRLTONYEAYSHLLCYLRGLNRQA-----TAEIRSLAHFCTSLQGLLSIAGVWATL 154
DB 86 RLRL--DAALAAALPPLLDVCRQAELNPRAPRLRLLEDAARQAARALGAARVALLAAL 143
QY 155 GY----PLQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWRSKDFNRL 205
DB 144 GAANRGPRAPPAATASA---ASATGVFPAKVLGLRVCGLYREWLSRTEGDLGQL 195

RESULT 10
US-09-901-257-3
; Sequence 3, Application US/09901257
; Patent No. US20020146707A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
; APPLICANT: Lawrence, David A.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret A.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Cardiotrophin-1 Compositions and Methods for the
; TITLE OF INVENTION: Treatment of Tumor
; FILE REFERENCE: P2533D3
; CURRENT APPLICATION NUMBER: US/09/901,257
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 09/648,252
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: US 09/033,114
; PRIOR FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: US 08/733,850
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: US 08/443,129
; PRIOR FILING DATE: 1995-05-17
; PRIOR APPLICATION NUMBER: US 08/286,304
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: US 08/233,609

; PRIOR FILING DATE: 1994-04-25
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Human
; US-09-901-257-3

Query Match 7.5%; Score 91.5; DB 10; Length 201;
Best Local Similarity 25.7%; Pred. No. 0.1;
Matches 45; Conservative 23; Mismatches 92; Indels 15; Gaps 5;

QY 40 IQKTYDITRYLEHQLRSLAGTYLNYLGGPNEPDPRLGAETLPRTATVNVLEWRSIND 99
DB 27 IROTHSLAHLTKYAEQLQYVQLQDPPGLPSFSPRLPVAGL-SAPAPSHAGLPVHE 85
QY 100 RLRLTONYEAYSHLLCYLRGLNRQA-----TAEIRSLAHFCTSLQGLLSIAGVWATL 154
DB 86 RLRL--DAALAAALPPLLDVCRQAELNPRAPRLRLLEDAARQAARALGAARVALLAAL 143
QY 155 GY----PLQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWRSKDFNRL 205
DB 144 GAANRGPRAPPAATASA---ASATGVFPAKVLGLRVCGLYREWLSRTEGDLGQL 195

RESULT 11
US-09-770-361-5
; Sequence 5, Application US/09770361
; Patent No. US20020123462A1
; GENERAL INFORMATION:
; APPLICANT: Fandl, James
; APPLICANT: Stahl, Neil
; TITLE OF INVENTION: Modified Ciliary Neurotrophic Factor, Method Of Making
; TITLE OF INVENTION: And Methods Of Use Thereof
; FILE REFERENCE: REG142
; CURRENT APPLICATION NUMBER: US/09/770,361
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 09/031,693
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 08/308,736
; PRIOR FILING DATE: 1994-09-19
; PRIOR APPLICATION NUMBER: 07/959,284
; PRIOR FILING DATE: 1992-10-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 195
; TYPE: PRT
; ORGANISM: CHICKEN
; US-09-770-361-5

Query Match 7.4%; Score 90; DB 10; Length 195;
Best Local Similarity 27.7%; Pred. No. 0.14;
Matches 52; Conservative 21; Mismatches 83; Indels 32; Gaps 10;

QY 46 LTRYLEHQLRSLAGTYLNYLGGPNEPDPRLGAETLPRTATVNVLEWRSINDRLTLQ 105
DB 23 LARKMRSDVTDLLDIYVERQG-----LDASISVAADVGVPTAAV--ERMAEQGTQRLLD 75
QY 106 N---YEAYSHLLCYLRGLNRQA---ATAELARSLA-----HFCTSLQGLLSIAGVMA 152
DB 76 NLAAYRAFRTLLAQMLEEQRELLGTDDELGPALAMLLQVSAFYVHLEELL-----ELE 130
QY 153 TIGYPLPQPLPGTEPAWAPGPAHSDFLQKMDDFWLKELQTLWRSKDFNRLKKMQPP 212
DB 131 SRGAPAE---GSEPP-AP-PRLSLFEQKLRLGLRLVRLAQWAVRSVRDLRLQSK--HGP 183
QY 213 AASVTIHL 220
DB 184 GSGAALGL 191

RESULT 12
US-10-000-776-2
; Sequence 2, Application US/10000776
; Patent No. US20020164609A1
; GENERAL INFORMATION:
; APPLICANT: Timans, Jacqueline C.
; APPLICANT: Pflanz, Stefan K.-H.
; APPLICANT: Kastelein, Robert A.
; APPLICANT: Bazan, Jose F.
; APPLICANT: Rennick, Donna
; APPLICANT: de Waal Malefyt, Rene
; APPLICANT: Cheung, Jeanne
; TITLE OF INVENTION: MAMMALIAN CYTOKINES; RELATED REAGENTS
; FILE REFERENCE: DX01040K3
; CURRENT APPLICATION NUMBER: US/10/000,776
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/791,497
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: 09/627,897
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/146,581
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/147,763
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 242
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-000-776-2

Query Match 7.1%; Score 87; DB 9; Length 242;
Best Local Similarity 22.6%; Pred. No. 0.36;
Matches 55; Conservative 26; Mismatches 106; Indels 56; Gaps 8;

QY 2 DLKAGDSWGMACLTCTVL-----WHLPAVPALNRTGDPGPGSIQK-----TYDLTR 48
DB 2 DLNPNKIGLSLLPLLLVQAGVWGFRPPG-----RPQLSLQELRREFTVSLHLAR 54
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRGAEITLPRATVNLVWRSNDLR 100
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRLSDP 102
QY 101 LRLTQNYEAYSHLLCVLRGLNRQAATAELRR-SLAHFTCTSLQGLGSLAGTYMATLGYPLP 159
DB 103 ERLCFISTTLQPPHAPLGLGLTQGRWTNMRMQLWAMRLDLRLQHLRFQVLAAGFNLP 162
QY 160 QP-----LPGTEPAWAGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204
DB 163 EEEEEEEEEERKGLLPGLGALGSAQVQSWPQLLSTYRLHLSLELVLSRAVRELL 222
QY 205 LKK 207
DB 223 LSK 225

RESULT 13
US-09-791-497-2
; Sequence 2, Application US/09791497
; Publication No. US20030008343A1
; GENERAL INFORMATION:
; APPLICANT: Timans, Jacqueline C.
; APPLICANT: Kastelein, Robert A.
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Pflanz, Stefan
; TITLE OF INVENTION: Mammalian Cytokines; Related Reagents
; FILE REFERENCE: DX01040K2
; CURRENT APPLICATION NUMBER: US/09/791,497
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: 09/627,897
; PRIOR FILING DATE: 2000-07-27

; PRIOR APPLICATION NUMBER: 60/146,581
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/147,763
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 242
; TYPE: PR1
; ORGANISM: primate; surmised Homo sapiens
US-09-791-497-2

Query Match 7.1%; Score 87; DB 9; Length 242;
Best Local Similarity 22.6%; Pred. No. 0.36;
Matches 55; Conservative 26; Mismatches 106; Indels 56; Gaps 8;

QY 2 DLKAGDSWGMACLTCTVL-----WHLPAVPALNRTGDPGPGSIQK-----TYDLTR 48
DB 2 DLNPNKIGLSLLPLLLVQAGVWGFRPPG-----RPQLSLQELRREFTVSLHLAR 54
QY 49 YLEHQLRSLAGTYLN-----YLGPPFNEPDPNPRGAEITLPRATVNLVWRSNDLR 100
DB 55 KLLSEVRGQAHRAESHLPGVNLVLLP-----LG-EQLPDVSLTFQAWRLSDP 102
QY 101 LRLTQNYEAYSHLLCVLRGLNRQAATAELRR-SLAHFTCTSLQGLGSLAGTYMATLGYPLP 159
DB 103 ERLCFISTTLQPPHAPLGLGLTQGRWTNMRMQLWAMRLDLRLQHLRFQVLAAGFNLP 162
QY 160 QP-----LPGTEPAWAGPAHSDFLQKMDDFWLLKELQTLWLRSAKDFNR 204
DB 163 EEEEEEEEEERKGLLPGLGALGSAQVQSWPQLLSTYRLHLSLELVLSRAVRELL 222
QY 205 LKK 207
DB 223 LSK 225

RESULT 14
US-09-810-052-2
; Sequence 2, Application US/09810052
; Patent No. US20020009775A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51
; FILE REFERENCE: 00-24
; CURRENT APPLICATION NUMBER: US/09/810,052
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/190,410
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/199,443
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 232
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-810-052-2

Query Match 7.0%; Score 86; DB 10; Length 232;
Best Local Similarity 22.7%; Pred. No. 0.42;
Matches 50; Conservative 25; Mismatches 95; Indels 50; Gaps 7;

QY 19 LWHLPAVPALNRTGDPGPGSIQK-----TYDLTRVLEHQLRSLAGTYLN----- 63
DB 15 VWGFRPPG-----RPQLSLQELRREFTVSLHLARLLEVRGQAHRAESHLPGVNL 67
QY 64 YLGPPFNEPDPNPRGAEITLPRATVNLVWRSNDLRRLRLTONYEAYSHLLCVLRGLNRQ 123
DB 68 YLLP-----LG-EQLPDVSLTFQAWRLSDPRLCFISTTLQPPHAPLGLGLTQ 115
QY 124 AATAELRR-SLAHFTCTSLQGLGSLAGTYMATLGYPLPOP-----LPGTEP 167

```
Db 116 GRWTNMRMQLWAMRLDLRLQHLRLQVLAAGFNLPDEEEEEEEEEERKGLLPGALG 175
Qy 168 AWAGCPAHSDFLQKMDDFWLLKELQETWLRSAKDFNLKK 207
Db 176 SALQGPQVSWPQLSTYRLHLSLELVLSRAVRELLLSK 215

RESULT 15
US-09-730-617-2
; Sequence 2, Application US/09730617
; Patent No. US20020068279A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E
; APPLICANT: Prayaga, Sudhirdas K
; APPLICANT: Shimkets, Richard A
; APPLICANT: Rastelli, Luca
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Mezes, Peter S
; TITLE OF INVENTION: No. US20020068279A1el Proteins and Nucleic Acids Encoding the Sam
; FILE REFERENCE: 15966-609
; CURRENT APPLICATION NUMBER: US/09/730,617
; PRIOR FILING DATE: 2000-12-05
; PRIOR APPLICATION NUMBER: 60/169,056
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 60/169,886
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/169,866
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/170,252
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 60/175,740
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-730-617-2

Query Match 6.7%; Score 81.5; DB 10; Length 348;
Best Local Similarity 28.2%; Pred. No. 1.9;
Matches 74; Conservative 19; Mismatches 88; Indels 81; Gaps 17;

Qy 9 WGMGLACLTVLWHLPAVPAALNRTGPGCPSP-----IQKTYDLTRYLHQHLSLAGT 60
Db 3 WSCLA--CTLLRVLPHVLSLUR--DFVPVPTGTLFHSCTSTNPCASFLEVAVEA-AGI 57
Qy 61 YLNYLGL--PP-----FNEPDP---FNPRLGAFTLPRATVNLVWRS--LN 98
Db 58 TPWTVGSEHPCCPYPSLHASPTDSFNRPSPAPLNRPISAGE--PRT---EAFPSGLK 111
Qy 99 DRLRLTQNYEAYSHLLCYLRLGNRQAATAELRRSLAHFCTSLQGLLGSAGWATLGYPL 158
Db 112 ARVGGTILAEA-----GLNSQGHAVEPVPS-----GPSGSGKGVLIKGRPS 153
Qy 159 PQPLP-----GTEPAWCPAHSDFLQKMDDFWLLK-----ELQTLWLRSAKD 201
Db 154 RMPKARECPVDRENLLLTNPVPSLLQLLSSSPFCIKVETEQRSAEFDLQS---RAARD 210
Qy 202 FN-RLKKKM-QPPAASVTIHL 221
Db 211 YNSRLLLKLGQIPAAKGSFLE 232
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:49:37 ; Search time 23.5426 Seconds
(without alignments)
6508.980 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagcttcgcggagcc.....tctctcttcgtctcccccc 797

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO spool/US09931704/runat 27012003 154126 3648/app query.fasta 1.7189
-DB=PIR_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704 @CGN 1 1.351 @runat 27012003 154126 3648 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Description
1	165	11.3	1532 2 A61262 collagen alpha 1(X
2	161	11.0	623 2 A45137 collagen alpha 4(I
3	160	11.0	383 2 S32975 BHLF1 protein
4	159	11.0	660 1 Q08E3 BHLF1 protein - hu
5	158	10.8	574 2 T32556 Wiskott-Aldrich sy
6	158	10.8	574 2 T38819 Wiskott-Aldrich sy
7	157	10.8	1843 2 S18803 collagen alpha 1(V
8	156.5	10.7	325 2 T32248 hypothetical prote
9	156	10.7	1690 1 CGHU1B collagen alpha 4(I
10	155.5	10.7	319 2 T32250 hypothetical prote
11	154.5	10.6	453 2 S18804 collagen alpha 4(I
12	152	10.4	382 2 S20375 collagen alpha 3(V
13	152	10.5	627 2 A44112 spidroin 2, dragli
14	151.5	10.4	1838 1 CGHU1V collagen alpha 1(V

15	150.5	10.3	302 2 T15936 hypothetical prote
16	150.5	10.3	418 2 T15142 hypothetical prote
17	150	10.3	278 2 B39066 proline-rich prote
18	149.5	10.3	1433 2 A46053 bullous pemphigoid
19	149	10.2	1146 2 A38587 collagen, cornea-s
20	148.5	10.2	1670 1 CGHU3B collagen alpha 3(I
21	147.5	10.1	317 2 T19143 hypothetical prote
22	147.5	10.1	684 2 A53019 collagen alpha 1(X
23	146	10.0	688 2 A53330 collagen alpha 2(I
24	145.5	10.0	304 2 T22602 hypothetical prote
25	145.5	10.0	317 2 A28996 proline-rich prote
26	144.5	9.9	240 2 A24264 proline-rich prote
27	144.5	10.0	385 2 T20410 hypothetical prote
28	144.5	9.9	1137 2 A86335 T20H2.9 protein -
29	144	9.9	1466 1 CGHU7L collagen alpha 1(I
30	143.5	9.8	317 2 T29960 hypothetical prote
31	143.5	9.8	435 2 T15143 hypothetical prote
32	143.5	9.9	481 2 A35628 loricerin - mouse
33	143	9.9	255 2 JQ0320 atrophin-1 related
34	143	9.8	1006 2 T42731 atrophin-1 related
35	143	9.8	1049 1 CGBO7S collagen alpha 1(I
36	143	9.8	3149 1 Q08E8 BPLF1 protein - hu
37	142.5	9.8	240 2 B24264 proline-rich prote
38	142.5	9.8	505 2 S72273 actin-depolymerizi
39	142	9.7	286 2 S34665 collagen, cuticula
40	142	9.7	707 2 A46302 PTB-associated spl
41	141.5	9.7	227 2 C29149 proline-rich prote
42	141.5	9.7	403 2 S52796 prpl2 protein - hu
43	141.5	9.8	1215 2 T32734 myosin-1A - Acanth
44	141	9.7	1201 2 G86441 unknown protein [i
45	140	9.7	839 2 F75518 hypothetical prote

ALIGNMENTS

RESULT 1

A61262
collagen alpha 1(XVII) chain - human (fragment)
N:Alternate names: bullous pemphigoid 180K autoantigen BPAG2; bullous pemphigoid antigen
C:Species: Homo sapiens (man)
C:Date: 12-May-1994 #sequence revision 12-Jul-1996 #text_change 21-Jul-2000
C:Accession: I56325; I55345; A61262
R:Giudice, G.J.; Emery, D.J.; Diaz, L.A.
J. Invest. Dermatol. 99, 243-250, 1992
A:Title: Cloning and primary structural analysis of the Bullous pemphigoid autoantigen
A:Reference number: I56325; MUID:92381323; PMID:1324962
A:Accession: I56325
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1532 <RES>
A:Cross-references: GB:M91669; NID:g179516; PIDN:AAA35605.1; PID:g179517
R:Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mattei, M.G.; Chu, M.L.; Uitto, J.
J. Biol. Chem. 266, 24064-24069, 1991
A:Title: Genomic organization of collagenous domains and chromosomal assignment of human
A:Reference number: I55345; MUID:92084712; PMID:1748679
A:Accession: I55345
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 543-890, 'P' <RE2>
A:Cross-references: GB:M63730; NID:g179520; PIDN:AAA51839.1; PID:g179521
R:Giudice, G.J.; Squigera, H.L.; Elias, P.M.; Diaz, L.A.
J. Clin. Invest. 87, 734-738, 1991
A:Title: Identification of two collagen domains within the bullous pemphigoid autoantigen
A:Reference number: A61262; MUID:91123476; PMID:1846881
A:Accession: A61262
A:Molecule type: mRNA
A:Residues: 543-890, 'P' <GIU>
C:Genetics:
A:Gene: GDB:COL17A1; BPAG2; BP180
A:Cross-references: GDB:I31396; OMIM:113811
A:Map position: 10q24.3-10q24.3

Alignment Scores:

Pred. No.: 0.000125 Length: 1532
Score: 165.00 Matches: 77
Percent Similarity: 29.41% Conservative: 23
Best Local Similarity: 22.65% Mismatches: 96
Query Match: 11.32% Indels: 144
DB: 2 Gaps: 13

US-09-931-704-1 (1-797) x A61262 (1-1532)

```
QY 20 CCGGGTCGCCCTCCCACTCCGACGCTCCGGAGAGGAGCCGACCCGCGGCCAG 79
Db 667 ArgGlyGluAlaGlyProProGlySerGlyGlyGluArgGlyAlaGlyGlu 686
QY 80 CCGACCCCATGACCTCCGAGC----- 103
Db 687 ProGlyProHisGlyProProGlyValProGlySerValGlyProLysGlySerSerGly 706
QY 104 -----AGGGACCTCGTGGGGATGTT 124
Db 707 SerProGlyProGlnGlyProProGlyProValGlyLeuGlnGlyLeuArgGlyGluVal 726
QY 125 AGCGTCGCT-----GTGACGGTCTCTGCGACCTCCCTGCGAGTCCGAGTCTCT 172
Db 727 GlyLeuProGlyValLysGlyAspLysGlyProMetGlyProProGlyProLysGlyAsp 746
QY 173 CAATCGCACAGGGACCCAGGCGCTGCCCTCCATCCAGAAACCTATGACCTCACCCG 232
Db 747 GlnGlyGluLysGlyProArgGlyLeuThrGlyGluProGly-----MetArgGly 763
QY 233 CTACCTGGAGCACAACTCCGCACTTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCCC 292
Db 764 LeuProGlyAla-----ValGlyGluPro 771
QY 293 TTTCACGAGCCAGACTTCAACCTCCCGCTGGGGGCGAGAGACTTGCCCGAGGCCAC 352
Db 772 GlyAlaLysGlyAlaMetGlyProAlaGlyProAspGlyHisGlnGlyProArgGlyGlu 791
QY 353 TGTGTACTGGAGGTGGCGAGCCTCAATGACAACTGCGGCTGACCCAGAACTACGA 412
Db 792 GlnGlyLeuThrGlyMet-----ProGlyIleArg 801
QY 413 GGCTACAGCACCTTCTGTACTTGTGCTGGCTCAACCGTCAGGCTGCCACTGTGA 472
Db 802 Gly-----ProGly-----ProSerGly----- 808
QY 473 GCTGCGCGGAGCTGC---CCACTTCTGCACGAGCTCCAGGGCTGCTGG--- 523
Db 809 -----AspProGlyLysProGlyLeuThrGlyProGlnGlyProGlnGlyLeuPro 825
QY 524 CAGCATTTGCGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCGCAGCGCTGCCCTGGAC 583
Db 826 GlyThrProGlyArgProGlyIleLysGlyGluProGlyAlaPro----- 840
QY 584 TGAACCCACTTGGACTCTGCGCCCTGCCCCACAGTACTTCTCCAGAGATGGAGACTT 643
Db 841 -----GlyLysIle 843
QY 644 CTGGCTGCTAAGAGCTGCAGACTGGCTGTG-----CGG 679
Db 844 ValThrSerGluGlySerSerMetLeuThrValProGlyProProGlyProProGlyAla 863
QY 680 CTGGCCCAAGACTTCAACCGCTCAAGAGAGATGAGCCTCCAGCAGCTGC----- 733
Db 864 MetGlyProProGlyProProGlyAlaProGlyProAlaGlyProAlaGlyLeuProGly 883
QY 733 ----- 733
Db 884 HisGlnGluValLeuAsnLeuGlnGlyProProGlyProProGlyProArgGlyProPro 903
QY 734 -----AGT 736
Db 904 GlyProSerIleProGlyProProGlyProProArgGlyProProGlyGluGlyLeuProGly 923
```

QY 737 CACCTGACCTGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCGTCCGCC 796
Db 924 ProProGlyProProGlySerPheLeuSerAsnSerGluThrPheLeuPheGlyProPro 943

RESULT 2

A45137
collagen alpha 4(IV) chain - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999
C:Accession: A45137
R:Kanagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain
A:Reference number: S28777; MUID:93054733; PMID:1429714
A:Accession: A45137
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-623 <KAM>
A:Experimental source: Basement membrane
A:Note: Sequence extracted from NCBI backbone (NCBIP:118549)
C:Superfamily: collagen alpha 1(IV) chain

Alignment Scores:
Pred. No.: 0.000246 Length: 523
Score: 161.00 Matches: 74
Percent Similarity: 36.51% Conservative: 14
Best Local Similarity: 30.71% Mismatches: 75
Query Match: 11.04% Indels: 78
DB: 2 Gaps: 16

US-09-931-704-1 (1-797) x A45137 (1-623)

```
QY 19 CCGGGCTCGCCCTCCCACTCCGCGAGGAGGAG-----CCGACCCCGGCC 72
Db 183 ProLysGlyAspIleProAspProGlyProGlyAspGlnGlyLeuProGlyProAsp 202
QY 73 GCGCCAGCCCGAGCCCATGGACCTCCGAGC-----AGGGG 108
Db 203 GlyProAspGlyAlaPro-GlyProAlaGlyProProGlySerValAspLeuLysG 222
QY 109 ACTCGTGGGGATGTTAGCGTGCCTGTGACGCTGCTCGGCACCTCCC----- 157
Db 222 yGluProGlyAspCysGlyValPro-----GlyProGlyProProGlyProProG 240
QY 158 -----TGCAGTGCC-----AGCTCTCAATCGCACAGGAGGCC----- 192
Db 240 yProProGlyCysGlnGlyValProGlyCysAspGlyHisAspGlyGlnLysGlyProMe 260
QY 193 GCGCTGGGCC---CTCCATCCAGAAACCTATGACCTACCCGCTACCTGGAGCACCAC 249
Db 260 tGlyPheProGlyLeuGlnGly-----ProHisGlyLeuProGlyLeuProG 276
QY 250 TCCGAGCTTGGCTGGGACCTATCTGAACTACTGTGGGCCCCCTTTCAACGAGCC----- 304
Db 276 yGluLys-----GlyLeuProGlySerProGlyArgLysGlyProTh 290
QY 305 -----AGACTTCAACCTCCCGCTGGGGCGAGAGACTTCTGCCAGGG 348
Db 290 xGlyProProGlyTyArgGlyGluProGlyProProAlaAspGluAspSerCysProAr 310
QY 349 CCACTGTTGACTTGGAGGTGCGGAAGCTCAATGACAACTGCGGCTGACCCAGAACT 408
Db 310 fileProGlyLeuProGlyValProGlyPro----- 320
QY 409 AGGAGGCTACAGCCACCTTCTGTGTTACTTGGTGGCCT-----CAACCGTCAGG 459
Db 321 -ArgGly-----ProGluGlyAlaMetGlyPheProGlyGlnArgGlyProProG 337
QY 460 CTGCCACTGCTGAGTCGCGCCG-----CAGCTGG 489
Db 337 yGlnGlyCysLysGlyGluProGlyLeuAspGlyLysArgGlyArgAspGlyValProG 357
QY 490 CCCACTTCTGCACCCAGCTCCAGGCGCTGCTGGGAGCATTTGGCGGCTCATGGCAGCTC 549
```


QY 540 TGACGCCCGCAATCTGCTCCACAGAGCCCTCGAGGCTGTGTCAGAGTGGG----- 490
 Db 246 -----GlnArgCysProAlaGlyProProPro---ThrArgSerGlyAlaAlaAala 261
 QY 489 -----CCAGGCTGC-----GGCGCAGCT 472
 Db 262 GlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysPro 281
 QY 471 CACGACTGGCAG-----CTGACGGTGTAGGCCAC 442
 Db 282 ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGln 301
 QY 441 GCAAGTAACACAGAGGTGCTGTAGGCTCTGTAGT----- 406
 Db 302 ArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAla 321
 QY 405 -----TCTGGTCAGCGGCGAGTTGTTCATTGAGGCTTCGCCACACCTCCAGT 358
 Db 322 AlaProGlyProGlyGlyAlaAlaVal-ProSerGlyAlaThrProHisProGluAr 341
 QY 357 CAACAGTGGCCCTGGGCAGAGTCTCTGCCCCAGCGGGGAGGCTGAAGTCTGGCTCGT 298
 Db 341 gGlySerGlyProAlaAspPro-----ProAlaAlaAlaArgLeuProProGluArgL 359
 QY 297 TGAA-----AGGGGGGCCAGGTAGTTCAGATAGTCCACGCCA 259
 Db 359 nGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyPro--Pro 378
 QY 258 AGCTGGGAGTTGCTGCTCAGGTAGCGGTGAGGTTCATAGTTTCTGGATGAGGGC 199
 Db 379 ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArg 398
 QY 198 CAGGCCCTGGGTCCTGTCGATTGAGAGCTGGCACTGGCAGGAGGTGCCAGAGCACCG 139
 Db 399 SerAlaArgAsnProGlyCysPro-ArgThrTrpArgArgSer----- 413
 QY 138 TGCACAGGCACGTAAATATCCCCACAGTCCCCTGCTCGAGAGTCCATGGGCTGGGGC 79
 Db 414 -GlyAlaGlnArgGlyHisProPro-ProGlyAlaGlyGlnArgProSerGlyProThrG 433
 QY 78 TGGGCGGG-----CCGGTGGGCTCTCTCCCGAGGCTGGCGAGTGGGAGTGGGAGGC 28
 Db 433 lyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyG 453
 QY 27 GAGCCCGCGCTCCGGCG 11
 Db 453 lyAlaAlaValProSer 458

RESULT 5
 T43556
 Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T43556
 R:Zankel, T.C.; Ow, D.W.
 Submitted to the EMBL Data Library, December 1997
 A:Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pombe,
 A:Reference number: 222575
 A:Accession: T43556
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-574 <ZAN>
 A:Cross-references: EMBL:AF038575; PIDN:AAB92587.1
 A:Experimental source: strain JS21
 C:Genetics:
 A:Gene: wapl
 A:Map position: I
 A:Introns: 72/3; 519/3; 564/1
 Alignment Scores:
 Pred. No.: 0.000412 Length: 574
 Score: 158.00 Matches: 77
 Percent Similarity: 34.74% Conservative: 22

Best Local Similarity: 27.02% Mismatches: 97
 Query Match: 10.84% Indels: 90
 DB: 2 Gaps: 13
 US-09-931-704-1 (1-797) x T43556 (1-574)
 QY 11 GCCTGAGCCGCGCTCCGCC-----TCCACTCCGCCACGCTCCGGGAG 55
 Db 319 ArgArgAsnArgGlyLysProProIleGlyAsnGlySerSerAsnSerSerLeu----- 336
 QY 56 AGGAGCCGACCCCGCCGCCAGCCAGCCCTGGACCTCCGAGCAGGGGACTCGTG 115
 Db 337 -----ProProProProProProProProProProProProProProProPro 348
 QY 116 GGGGATGTAGCGTCTGTGCACGGTGTCTGGCA-----CCTCCTCGAGT 163
 Db 349 GlySerIleProLeuProProGlnArgSerAlaProProProProProProProPro 368
 QY 164 GCCAGCTCTCAATCCACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAAACCTATGA 223
 Db 369 AlaProSerThrGlyArgGlnProPro-----ProLeuSerSerSerArgAlaVal 385
 QY 224 CTCACCCG-----CTACCTGGAGCACCACTCCGAGCTTGCTGG 265
 Db 386 SerAsnProProAlaProProProAlaIleProGlyArgSerAlaProAlaLeu----- 403
 QY 266 GACCTATCTGAATACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCCT 325
 Db 404 -----ProProLeuGlyAsnAlaSerArgThrSerThrProPro 416
 QY 326 GGGGCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGA 385
 Db 417 -----ValProThrProProSerLeu 423
 QY 386 CAACCTGGCTGACCCAGAACTACAGGCCTACAGCCACCTCTGTGTACTTGGTGG 445
 Db 424 ProProSerAlaProProSerLeuProProSerAlaProProSerLeuProMetGlyAla 443
 QY 446 CCTAACCGTCAGGCTGCCACTGTCTGAGCTGGCGCGCAGCTGGCCCTTCTGCACCAG 505
 Db 444 ProAla-----AlaProProLeuProProSerAlaProAla 456
 QY 506 CTCCAGGCTCTGTGGCAGCATTTGGGGGCTCATGGCAGCTCTGGCTACCCACTGCC 565
 Db 457 ProProLeuProAla-----GlyMetProAlaAla 466
 QY 566 CAGCGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCACAGTCA----- 619
 Db 467 Pro-ProLeuProProAlaAlaProAlaProProProProProProProProAla 486
 QY 620 -----CTTCTCCAGAAATGGACGACTT----- 643
 Db 486 aProValAlaSerIleAlaGluLeuProGlnGlnAspGlyArgAlaAsnLeuMetAlaSe 506
 QY 644 -CTGGCTCTGAAGGAGTGCAGACTGGCTGTGGCGCTTCGGCCAAAGGACTTCAACCGGC 702
 Db 506 rIleArgAlaSerGlyGlyMetAspLeuLeuLysSerArgLysValSerAlaSerProSe 526
 QY 703 TCAGAGAAGATGAGGCTCCAGCAGCTGCAGTCCCT-----GCACCTGGGG 753
 Db 526 r-----ValAlaSerThrLysThrSerAsnProProValGluAlaProProSe 542
 QY 754 CTCATGGCTCTG 766
 Db 542 rAsnAsnLeuMet 546

RESULT 6

T38819

wiskott-aldrich syndrome protein homolog 1 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T38819
 R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

QY 454 GTGAGTGTCCACTGTGAGTGGCGG-----CAGCTGGCCCACTT-----CTGCACCA 504
Db 1186 oThrGlyProGln-----GlyProIleGlyGlnProGlyProSerGlyAlaAspGly 1203
QY 505 GCCTCCAGGCGCTGCTGGCGAG-----CATTCGGCGCTCATGGCAGCTC 549
Db 1203 yGluProGlyProArgGlyGlnGlyLeuPheGlyGlnLysGlyAspGlyGlySerAr 1223
QY 550 TGGCTTACCACCTGCCCCAGCGCTGCTGGGACTGAACCCACTTGGACTCTGGTGGCCCTG 609
Db 1223 gGlyPheProGlyProProGlyPro-----ValGlyLeuGlnGlyLeuProGly 1239
QY 610 CCCACAGTACTTCTCCAGAGATGGA 637
Db 1239 yPro-----ProGlyGluLysGly 1245
RESULT 8
T32248
hypothetical protein T15B7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32248
R:Pauley, A.; Gattung, S.
Submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T15B7.
A:Reference number: Z31139
A:Accession: T32248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-325 <PAU>
A:Cross-references: EMBL:AF022985; PIDN: AAB69961.1; GSPDB: GN00023; CESP: T15B7.4
A:Experimental source: strain Bristol N2; clone T15B7
C:Genetics:
A:Gene: CESP: T15B7.4
A:Map position: 5
A:Introns: 266/1
C:Superfamily: unassigned collagens
Alignment Scores:
Pred. No.: 0.000529 Length: 325
Score: 156.50 Matches: 66
Percent Similarity: 35.19% Conservative: 10
Best Local Similarity: 30.56% Mismatches: 86
Query Match: 10.73% Indels: 54
DB: 2 Gaps: 10
US-09-931-704-1 (1-797) x T32248 (1-325)
QY 56 AGGAGCGCACC CGCGCGCCGACGCG-----CAGCCCCATGGACCTCCG 100
Db 85 GlnSerArgGlyCysProAlaGlyProGlyProGlyGlnProGlyAlaProGly 104
QY 101 AGCAGGGGACTCGTGGGGGATGTTAGCTGCTGTGCGACGGTGTCTGGCACCCTCC--- 157
Db 105 GluGlnGlyHisProGlyLeuAlaGlyGlnProGlySerGlyAlaArgIleAsnProAla 124
QY 158 -----TGCAGTGCAGCTCTCAATCGCACGGGACCC 190
Db 125 ThrGlyArgProGlyPheCysIleThrCysProAlaGlyAlaProGlyProAlaGlyPro 144
QY 191 AGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCCTGCTACCTGGAGCACCAACT 250
Db 145 -----ProGlyAlaProGly 149
QY 251 CCGCAGCTTGGTGGGACCTATCTGAACCTACTGGGCCCCCCTTCAACA-----GCCAGA 307
Db 150 ProLys-----GlyAsnAsnGlyGlnProGlyAlaProAlaGlnSerGlyGlyArg 166
QY 308 CTTCAACCTCCCGCCTGG-----GGCAGAGACTCTGCCAGGCGCCAC 352
Db 167 GlyProProGlyProArgGlyProAlaGlyAspAlaGlySerProGlyGlnProGlyHis 186
QY 353 TGTGACTTGGAGGTGTGGCGAAGCCTCAATGATCAAACTGCGGCTGACCCAGAACTACGA 412

Db 187 -----ProGlySerProGlyAsnProGlyArgGlyGlyGlnArgSerArg 201
QY 413 GGCCTACAGCACCTTCTGTACTTGGTGGCTGCCTCAACCGTCAGGC----- 460
Db 202 GlyLeuProGlyProSerGlyArgProGlyProProGlyProAlaGlyGlyProGlyGln 221
QY 461 ---TGCCACTGTGAGCTCGCGCAGCTGGCCCACTTCTGCACACGCTCCAGGCGCT 517
Db 222 ProGlyHisSerGlyGlyAlaGlySerProGlyPro-----GlnGlyProProGlyPro 239
QY 518 GCTGGGCGAG-----CATTCGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCCGCC 571
Db 240 SerGlyGlnProGlyHisSerGlyAsnAspGlyValProGlyAlaProGlyAsnProGly 259
QY 572 GCTGCTGGGACTGAACCCACTTGGACTCTGGCGCTGCCACACAGT 617
Db 260 Ser-ProGlyGlyAspAlaAlaTyrCysProCysProAlaArgSer 274
RESULT 9
CGHUIB
collagen alpha 4(IV) chain precursor - human
N:Alternate names: procollagen alpha 4(IV) chain
C:Species: Homo sapiens (man)
C:Date: 06-Feb-1995 #sequence_revision 03-Oct-1995 #text_change 16-Jun-2000
C:Accession: A53360; S36854; S28777
R:Leinonen, A.; Maruyama, M.; Mochizuki, T.; Tryggvason, K.; Reiders, S. T.
J. Biol. Chem. 269, 26172-26177, 1994
A:Title: Complete primary structure of the human type IV collagen alpha4(IV) chain.
A:Reference number: A53360; MUID: 95014445; PMID: 7523402
A:Accession: A53360
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1690 <LEI>
A:Cross-references: GB:X81053; NID: G574805; PIDN: CRA56943.1; PID: G574806
R: Sugimoto, M.; Ohashi, T.; Yoshioka, H.; Matsuo, N.; Ninomiya, Y.
FEBS Lett. 330, 122-128, 1993
A:Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen
A:Reference number: S36854; MUID: 93374047; PMID: 8365481
A:Accession: S36854
A:Molecule type: DNA; mRNA
A:Residues: 1219-1658, 'FE', 1661-1690 <SUG>
A:Cross-references: DBJ: D17391; NID: G440365; PIDN: BAA04214.1; PID: G457161
R: Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain
A:Reference number: S28777; MUID: 93054733; PMID: 1429714
A:Accession: S28777
A:Molecule type: DNA
A:Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>
A:Cross-references: GB: L01475; GB: L01476
A:Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB: COL4A4
A:Cross-references: GDB: 132673; OMIM: 120131
A:Map position: 2q35-2q37
A:Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w
C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeri
er associations in the interrupted helical domain (with disulfide and desmosine cross
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidne
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydro
F; 1-38/Domain: signal sequence #status predicted <SIG>
F; 39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
F; 39-61/Domain: amino-terminal nonhelical, NHI <NHI>
F; 62-1466/Region: interrupted helical
F; 94-96/Region: cell attachment (R-G-D) motif

F:145-147/Region: cell attachment (R-G-D) motif
F:189-191/Region: cell attachment (R-G-D) motif
F:310-312/Region: cell attachment (R-G-D) motif
F:724-726/Region: cell attachment (R-G-D) motif
F:785-787/Region: cell attachment (R-G-D) motif
F:788-789/Region: cell attachment (R-G-D) motif
F:789-791/Region: cell attachment (R-G-D) motif
F:1212-1214/Region: cell attachment (R-G-D) motif
F:1467-1690/Domain: carboxyl-terminal nonhelical, N1 <N1>
F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1579-1686/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:147-52-55-57-266-400-492-494-668-790-828-1095-1131-1294-1317-1375-1407/Disulfide bond
F:146-665/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1480-1566/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
F:1525-1531,1634-1641/Disulfide bonds: #status predicted
F:1588-1683,1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted

Alignment Scores:
Pred. No.: 0.000595 Length: 1690
Score: 156.00 Matches: 73
Percent Similarity: 35.10% Conservative: 13
Best Local Similarity: 29.80% Mismatches: 82
Query Match: 10.70% Indels: 77
DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x CGHUB (1-1690)

QY 12 GCGGAGCGCGCTCGCCCT-----CCCATCCCGCAGCTCCGGGAGGGA 59
DB 1244 AlaGlyAlaThrGlyArgAlaProLysAspIleProAspPro-GlyProGlyAspGI 1263
QY 60 G-----CGCACCCGCGCGCCAGCCCGCCATGACCTCCGAGC----- 103
DB 1263 nGlyProProGlyProAspGlyProArgGlyAlaPro-GlyProGlyLeuProGly 1283
QY 104 -----AGGGACTCGTGGGGATGTAGCTGCTGCACGCTGCTCTGG 149
DB 1283 erValAspLeuLeuArgGlyGluProGlyAspCysGlyLeuPro-----GlyProProG 1301
QY 150 CACTCC-----TGCGAGTCCGAGCT 170
DB 1301 lProProGlyProProGlyProProGlyTyrlsGlyPheProGlyCysAsp---GlyL 1320
QY 171 CTCAATGCGACAGGAGCCAGGCGCTCCCTCCATCCAGAAACCTATGACCTCACC 230
DB 1320 yeAspGlyGlnLysGlyProMetGlyPhePro---GlyProGlnGly-----ProHig 1337
QY 231 CGTACTCGGACGACCAACTCCGAGCTTGGTGGGACCTATCTGAACCTACCTGGGCCCC 290
DB 1337 lPheProGlyProProGlyGluLys-----GlyLeuProGlyProP 1351
QY 291 CCTTTCAACGAGCC-----AGACTTCAACCTCCCGCCCTGGGG 329
DB 1351 roGlyArgLysGlyProThrGlyLeuProGlyProGlyGluProGlyProAlaA 1371
QY 330 GCAGAGACTTGCAGGCGCACTGTGACTTGGAGGTGGGAGACCTCAATGACAA 389
DB 1371 spValAspAspCysProArgIleProGlyLeuProGlyAla----- 1384
QY 390 CTGCGGCTGACCCAGCACTACGAGGCTCAGCCACCTCTGTGTTACTTGGCGCTC 449
DB 1385 -----ProGlyMetArgGlyProGluGlyAlaMetGlyLeuProGlyMetArg 1401
QY 450 AACCGTCAGGCTGCCACTGTGAGCTGCGCGC----- 481
DB 1401 lProProGlyProGlyCysLysGlyGluProGlyLeuAspGlyArgGlyValAspG 1421
QY 482 --CAGCTGCGCCACTTCTGSCACAGCCCTCCAGGCGCTGCTGGGACGATTCGGGCGTC 539
DB 1421 lValProGly-----SerProGlyProProGlyArgLys---GlyAspT 1435
QY 540 ATGGAGCTCTGGCTACCCACTGCGCCCGCCGCTGCTGGGACTGACCCACTTGGACT 599
DB 1435 hrGlyGluAspGlyTyrlsGlyProGlyProProGly-----ProIleGlyAsp 1452

QY 600 CCTGGCCCT 608
DB 1453 ProGlyPro 1455

RESULT 10

T32250
hypothetical protein T15B7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32250
R:Pauley, A.; Gattung, S.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T15B7.
A:Reference number: Z21139
A:Accession: T32250
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-319 <PAU>
A:Cross-references: EMBL:AF022985; PIDN:AAB69959.1; GSPDB:GN00023; CESP:T15B7.3
A:Experimental source: strain Bristol N2; clone T15B7
C:Genetics:
A:Gene: CESP:T15B7.3
A:Map position: 5
A:Introns: 266/1
C:Superfamily: unassigned collagens

Alignment Scores:
Pred. No.: 0.000629 Length: 319
Score: 155.50 Matches: 68
Percent Similarity: 35.43% Conservative: 11
Best Local Similarity: 30.49% Mismatches: 84
Query Match: 10.67% Indels: 60
DB: 2 Gaps: 11

US-09-931-704-1 (1-797) x T32250 (1-319)

QY 53 GAGAGGACCGGCGCGCCAGCCCGCCATGAGCCCGCAGCAGC----- 106
DB 84 GluGlnSerArgGlyCysProAlaGlyProPro-----GlyProGlyGlnProGly 101
QY 107 -----GGACTCGTGGGGATGTAGCGTCTGTCAGCGTCTCTGGCA 151
DB 102 AlaArgGlyGluAlaGlyLeuProGlyIleAlaGlyGlnProGlySerGlyAlaArgIle 121
QY 152 CCTCCC-----TGAGTCCGAGCTCTCAATCGCAC 181
DB 122 AsnProAlaThrGlyArgProGlyPheCysIleThrCysProAlaGlyAlaProGlyPro 141
QY 182 AGGGGACCGGCGCTTGGCGCTCCATCCAGAAACCTATGACCTCACCCGCTACCTGGA 241
DB 142 AlaGlyProProGlyAlaProGlyProLysGlyAsn-----AsnGlyGlnProGly 158
QY 242 GCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCGCCCTTCAACGA 301
DB 159 AlaProAlaGlnSerGlyGly-----ArgGlyProProGlyProArg 172
QY 302 GCCAGACT-----CAACCTCCCGCTGG----- 328
DB 173 GlyProAlaGlyAspAlaGlySerProGlyGlnProGlyHisProGlySerProGlyAsn 192
QY 329 ---GGCAGAGACTCTCCCGGCGCCACTGTGACTTGGAGGTGGCGAAGCCCTCAATGA 385
DB 193 ProGlyArgGlyGlyGlnArgSerArgGlyThrProGlyAlaSerGlyArgPro----- 210
QY 386 CAAACTGCGGCTGACCCAGAACTACAGGCGCTACAGCCACTTCTGTGTACTTGGCTGG 445
DB 211 -----GlyProGlnGlyProAlaGlyAlaProGlyGln 221
QY 446 CCTCAACCTCAGGCTGCCACTGCTGAGCTGGCGCCGAGCTCTTCTGCTGCTGCTGCTG 505
DB 222 ProGlyArgSerGly-----GlyAlaGlyThrProGlyPro-----GlnGly 235

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:44:12 ; Search time 10.3445 Seconds
(without alignments)
6391.168 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagcttcgagccgagcc.....tctctctctctgctcccccc 797

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cn2_1/USPTO_spool/US09931704/runat_27012003_154125_3623/app_query.fasta.1.7189
-DB=SwissProt_40 -QMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09931704 @CGN 1 1 141 @runat_27012003_154125_3623 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WAE TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	161	11.0	623	1 CA44_RABIT	P5787 oryctolagus
2	159	11.0	660	1 YHL1_EBV	P03181 epstein-bar
3	156	10.7	1690	1 CA44_HUMAN	P53420 homo sapien
4	155.5	10.7	2715	1 TRX2_HUMAN	Q9um66 homo sapien
5	154.5	10.6	453	1 CA44_BOVIN	Q29442 bos taurus
6	154.5	10.6	1838	1 CA15_HUMAN	P20908 homo sapien
7	154	10.6	1083	1 T2D3_HUMAN	O00268 homo sapien
8	152	10.4	382	1 CA35_HUMAN	P25940 homo sapien
9	152	10.5	627	1 SP02_NEPC	P46804 nephila cia
10	148.5	10.2	296	1 PRP3_MOUSE	P05143 mus musculus
11	148.5	10.2	1670	1 CA34_MOUSE	Q01955 homo sapien
12	147.5	10.1	234	1 PRPM_HUMAN	P10161 homo sapien
13	147.5	10.1	317	1 YQ35_CABEL	Q09456 caenorhabdi
14	147.5	10.1	555	1 GP1_CHLRE	Q9fpq6 chlamydomon
15	147	10.1	1183	1 DRPL_RAT	P54258 rattus norv
16	144.5	9.9	261	1 PRP2_MOUSE	P05142 mus musculus
17	144.5	9.9	534	1 APG_ARATH	P40602 arabidopsis
18	144	9.9	1466	1 CA13_HUMAN	P02461 homo sapien

C	19	143.5	9.9	481	1 LORI_MOUSE	P18165 mus musculus
	20	143	9.8	1049	1 CA13_BOVIN	P04258 bos taurus
	21	143	9.8	1336	1 W146_HUMAN	Q9cuj8 homo sapien
	22	143	9.8	3149	1 TEGU_EBV	P03186 epstein-bar
	23	142.5	9.8	505	1 WASL_BOVIN	Q95107 bos taurus
	24	142	9.7	707	1 SFPO_HUMAN	P23246 homo sapien
	25	140.5	9.6	2944	1 CA17_HUMAN	Q02388 homo sapien
	26	139.5	9.6	316	1 CC07_CABEL	P18832 caenorhabdi
	27	139.5	9.6	475	1 S3A2_MOUSE	O62203 mus musculus
	28	139.5	9.6	503	1 WAP1_HUMAN	O43516 homo sapien
	29	139.5	9.6	505	1 WASL_HUMAN	O00401 homo sapien
C	30	139.5	9.6	1147	1 MYSB_ACACA	P19706 acanthamoeb
	31	139.5	9.6	1669	1 CA14_MOUSE	P02463 mus musculus
	32	139	9.5	744	1 CA18_RABIT	P14282 oryctolagus
	33	138.5	9.5	497	1 HMB5_DROME	P18488 drosophila
	34	138.5	9.5	620	1 EXTN_TOBAC	P13983 nicotiana t
	35	138.5	9.5	743	1 CA18_MOUSE	Q00780 mus musculus
	36	138	9.5	276	1 PRPL_HUMAN	P10162 homo sapien
	37	138	9.5	497	1 WAS2_HUMAN	Q9Y6W5 homo sapien
	38	137.5	9.4	1460	1 CA11_CANFA	Q9X817 canis famil
	39	137.5	9.4	1464	1 CA11_HUMAN	P02452 homo sapien
	40	137	9.4	174	1 PRPP_HUMAN	P81489 homo sapien
	41	137	9.4	331	1 PRPL_HUMAN	P04280 homo sapien
	42	137	9.4	1516	1 CA1H_HUMAN	P39060 homo sapien
	43	137	9.4	1763	1 CA24_ASCSU	P27393 ascaris suu
	44	136.5	9.4	1527	1 CA1H_MOUSE	P39061 mus musculus
	45	136.5	9.4	1567	1 FMN2_MOUSE	Q9J104 mus musculus

ALIGNMENTS

RESULT 1
CA44_RABIT
ID AC P55787; STANDARD; PRT; 623 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 4(IV) chain (Fragment).
GN COL4A4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Corneal endothelium;
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE',
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; L01477; -; NOT ANNOTATED CDS.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001442; ProcollagnC4.
 DR Pfam; PF01391; Collagen; 5.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD003923; ProcollagnC4; 2.
 DR SMART; SM00111; C4; 2.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 DR NON_TER 1
 FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
 FT DOMAIN 393 623 NONHELICAL REGION (NC1).
 FT DISULFID 413 502 OR 499 (BY SIMILARITY).
 FT DISULFID 446 499 OR 502 (BY SIMILARITY).
 FT DISULFID 458 464 BY SIMILARITY.
 FT DISULFID 521 619 OR 616 (BY SIMILARITY).
 FT DISULFID 555 616 OR 619 (BY SIMILARITY).
 FT DISULFID 567 574 BY SIMILARITY.
 SQ SEQUENCE 623 AA; 62393 MW; CCBG9BB31242FB82 CRC64;

 Alignment Scores:
 Pred. No.: 0.00468 Length: 623
 Score: 161.00 Matches: 74
 Percent Similarity: 36.51% Conservative: 14
 Best Local Similarity: 30.71% Mismatches: 75
 Query Match: 11.04% Indels: 78
 DB: 1 Gaps: 16

 US-09-931-704-1 (1-797) x CA44_RABIT (1-623)
 QY 19 CCGGGCTCGCCCTCCCACTCCGCAGCTCCGGAGGAGAG-----CCGACCCCGGCC 72
 Db 183 ProlysGlyAspIleProAspProGlyProProGlyAspGlnGlyLeuProGlyProAsp 202
 QY 73 GGCCAGCCCGCAGCCCACTGACCTCCGAGC-----AGGGG 108
 Db 203 GlyProargGlyAlaPro-GlyProAlaGlyProProGlySerValaspLeuLeuLysG 222
 QY 109 ACTCGTGGGGATGTTAGCGTCTGTGTCACAGTGTCTTGGCACTCC----- 157
 Db 222 YGluProGlyAspCysGlyValPro-----GlyProProGlyProProGlyProProG 240
 QY 158 -----TGCAGTGCC-----AGCTCTCAATCGCAGGGGACCCAG 192
 Db 240 YProProGlyCysGlnGlyValProGlyProGlyCysAspGlyHisAspGlyGlnLysGlyProMe 260
 QY 193 GGCTCGGCC---CTCCATCCAGAAAACCTATGACCTCACCGCTTACCTGGAGCACCAC 249
 Db 260 tGlyPheProGlyLeuGlnGly-----ProHisGlyLeuProGlyLeuProG 276
 QY 250 TCCGACGCTTGGCTGGGACCTATCTGAATACCTGGGCCCCCTTTCAACAGGC----- 304
 Db 276 YGluLys-----GlyLeuProGlySerProGlyArgLysGlyProTh 290
 QY 305 -----AGACTTCAACCTCCCGCTCGGGCAGAGACTCTGCCAGGG 348
 Db 290 rGlyProProGlyTyArgGlyGluProGlyProProAlaAspGluAspSerCysProAr 310
 QY 349 CCACCTGTTGACTTGGAGGTGGGGAAGCCCTCAATGACAAACTGGCGCTGACCCAGAACT 408
 Db 310 gIleProGlyLeuProGlyValProGlyPro----- 320
 QY 409 ACGAGGCTACAGCCACCTTCTGTGTATTACTTTCGTGGCT-----CAACCGTCTAGG 459
 Db 409 ACGAGGCTACAGCCACCTTCTGTGTATTACTTTCGTGGCT-----CAACCGTCTAGG 459

Db 195 ProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisPro 214
 QY 633 TCTTCTGGAGGAAGTCACTGTGGG-----CAGGCGCCAG 601
 Db 215 GluArgGlySerGlyProAlaAlaProProAlaAlaAlaAlaAlaAlaAlaAlaAla 234
 QY 600 GAGTCAAGTGGTTCAGTCCAGGAGGCGGTGGGGCAGTGGGTAGCCAGAGCTGCCA 541
 Db 235 GluProArgLeuProGlnAspLeuAlaAlaAla----- 245
 QY 540 TGACGCCGCCCAATGCTGCCAGCAGGCGCTGGAGGCTGGTGCAGAGTGGG----- 490
 Db 246 -----GlnArgCysProAlaGlyProProPro-----ThrArgSerGlyAlaAlaAla 261
 QY 489 -----CCAGCTGC-----GGCGCAGCT 472
 Db 262 GlnArgThrHisArgArgProProGlyCysProArgSerAlaAlaArgAsnProGlyCysPro 281
 QY 471 CAGCAGTGGCAG-----CCTGACGGTTGAGGCCAC 442
 Db 282 ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGln 301
 QY 441 GCAAGTAACACAGAGGTGCTAGCGCTCGTAGT----- 406
 Db 302 ArgProSerGlyProThrGlyGlyArgProAlaAlaAlaProGlyAlaProGlyThrProAla 321
 QY 405 -----TCTGGGTGAGCGCGAGTTGTTCATTGAGGCTTCGCCACACCTCCAACT 358
 Db 322 AlaProGlyProGlyGlyAlaAlaVal-ProSerGlyAlaThrProHisProGluArg 341
 QY 357 CAACAGTGGCTGGCGCAGAGTCTCTGCCCGAGGCGGAGGTTGAAGTCTGGCTCGT 298
 Db 341 gGlySerGlyProAlaAlaPro-----ProAlaAlaAlaAlaArgLeuProProGluArgG 359
 QY 297 TGAA-----AGGGGGCCCGAGGTAGTTCAGATAGTCCCGAGCCA 259
 Db 359 nGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyPro--Pro 378
 QY 258 AGCTGCGAGTGGTCTCCAGGTAGCGGTGAGGTTCATAGTGTTCGTGATGGAGGGC 199
 Db 379 ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArg 398
 QY 198 CAGGCGCTGGTCCCTGTGCGATTGAGAGTGGCTGCGAGGAGTCCAGAGCAGCCG 139
 Db 399 SerAlaArgAsnProGlyCysPro-ArgThrTrpArgArgSer----- 413
 QY 138 TGACAGGCGAGTCAATCCCGAGTCCCTCTCGAGGTCCAGTGGGCTGGGC 79
 Db 414 -GlyAlaGlnArgGlyHisProPro-ProGlyAlaGlyGlnArgProSerGlyProThrG 433
 QY 78 TGGGCGCG-----CCGGGTGCGGCTCTCTCCGAGGCTGGCGAGTGGGAGGGC 28
 Db 433 lGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyG 453
 QY 27 GAGCGCGCTCGGCG 11
 Db 453 lAlaAlaValProSer 458
 RESULT 3
 CA44 HUMAN
 ID CA44 HUMAN STANDARD; PRT; 1690 AA.
 AC P53420;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 4(IV) chain precursor.
 OS COL4A4.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;
 RX MEDLINE=95014445; PubMed=7523402;
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reiders S.T.;
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)
 chain. Comparison with structure and expression of the other alpha
 (IV) chains.";
 RL J. Biol. Chem. 269:26172-26177 (1994).
 RN [2]
 RP SEQUENCE OF 1-23 FROM N.A.
 RX MEDLINE=98196854; PubMed=9537506;
 RA Momota R., Sugimoto M., Ohashi T., Kigawa K., Yoshioka H.,
 RA Ninomiya Y.;
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
 alpha4(IV) collagen chains are arranged head-to-head on chromosome
 2q36.";
 RL FEBS Lett. 424:11-16 (1998).
 RN [3]
 RP SEQUENCE OF 1219-1690 FROM N.A.
 RX TISSUE=Eye;
 MEDLINE=93374047; PubMed=8365481;
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
 RT "cDNA isolation and partial gene structure of the human alpha 4(IV)
 collagen chain.";
 RL FEBS Lett. 330:122-128 (1993).
 RN [4]
 RP SEQUENCE OF 1407-1507 FROM N.A.
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 alpha 4 chain of basement membrane collagen type IV and assignment of
 the gene to the distal long arm of human chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758 (1992).
 RN [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=9738662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499 (1997).
 RN [6]
 RP VARIANT AS SER-1201.
 RX MEDLINE=95078927; PubMed=7987396;
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
 RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
 RA Smeets H.J.M., Reiders S.T.;
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
 collagen genes in autosomal recessive Alport syndrome.";
 RL Nat. Genet. 8:77-82 (1994).
 RN [7]
 RP VARIANT BFH GLU-897.
 RX MEDLINE=96379660; PubMed=8787673;
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
 RT "Benign familial hematuria due to mutation of the type IV collagen
 alpha4 gene.";
 RL J. Clin. Invest. 98:1114-1118 (1996).
 RN [8]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99011253; PubMed=9792860;
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
 RA Cochot P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
 RT "Determination of the genomic structure of the COL4A4 gene and of
 novel mutations causing autosomal recessive Alport syndrome.";
 RL Am. J. Hum. Genet. 63:1329-1340 (1998).
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
 CC NIDOGEN.
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,

Db 1401 lyProGlyProGlyCysLysGlyGluProGlyLeuAspGlyArgGlyValAspG 1421
 Qy 482 --CACCTGGCCCACTTTCACACAGCTCCAGGCGCTGCTGGGCGAGCATTCGGGGCGTC 539
 Db 1421 lyValProGly-----SerProGlyProGlyArgLys---GlyAspT 1435
 Qy 540 ATGGCAGCTCTGGCTACCACTGCCAGCGCTGCTGGGAGCTGAACCCACTTGGACT 599
 Db 1435 hrGlyGluAspGlyTyrProGlyGlyProGlyPro-ProGly-----ProIleGlyAsp 1452
 Qy 600 CTGGCCCT 608
 Db 1453 ProGlyPro 1455
 RESULT 4
 TRX2 HUMAN
 ID TRX2 HUMAN STANDARD; PRT; 2715 AA.
 AC Q9UMN6; Q9UK25; Q9Y666; Q9Y668; Q15022; Q96GP2; Q96IP3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trithorax homolog 2 (Mixed lineage leukemia gene homolog 2 protein).
 GN TRX2 OR HRX2 OR MLL2 OR MLL4 OR KIAA0304.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Angrand P.O., Valvaine H., Jeanmougin F., Adamson A.,
 van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
 Lamerding J., Chambon P., Lessou R., Stewart A., Asaland R.;
 RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin
 RT regulators which contain PHD fingers and SET domains";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Lamerding J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,
 Garcia E., Kyle A., Ramirez M., Stilwagen S., Garnes J., Danganan L.,
 Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
 Carrano A.V.;
 RT "Sequence analysis of a 1 Mb region in human 19q13.1";
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).
 RC TISSUE=Testis, and Leukocyte;
 EX MEDLINE=20105772; PubMed=10637508;
 RA Huntsman D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,
 Wiedemann L.M., Aparicio S., Caldas C.;
 RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps
 RT to 19q13.1 and is amplified in solid tumor cell lines";
 RL Oncogene 18:7975-7984(1999).
 RN [4]
 RP SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 Miyaajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro";
 RL DNA Res. 4:141-150(1997).
 RN [5]
 RP SEQUENCE OF 1918-2715 FROM N.A.
 RC TISSUE=Brain, and Skin;
 RA Strausberg R.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).
 RC TISSUE=Placenta, and Bone marrow;
 RX MEDLINE=99339983; PubMed=10409430;
 RA Fitzgerald K.T., Diaz M.O.;
 RT "MLL2: A new mammalian member of the trx/MLL family of genes.";

Genomics 59:187-192(1999).
 CC -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.
 CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,
 CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL
 CC BLOOD LYMPHOCYTES, AND PLACENTA.
 CC -!- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.
 CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -!- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC -!- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ007041; CAB45385.1; -;
 DR EMBL; AD000671; -; NOT ANNOTATED_CDS.
 DR EMBL; AB002302; BAA20763.2; -;
 DR EMBL; AF186605; AAD56420.1; -;
 DR EMBL; AF104918; AAD17932.1; -;
 DR EMBL; AF105279; AAD26113.1; -;
 DR EMBL; BC009337; AAH09337.1; -;
 DR EMBL; BC007353; AAH07353.1; -;
 DR EMBL; AF105280; AAD26112.1; -;
 DR MIM; 606834; -;
 DR InterPro; IPR000637; AT hook.
 DR InterPro; IPR003889; FYRICH_C.
 DR InterPro; IPR003888; FYRICH_N.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR002857; Znf CXXC.
 DR InterPro; IPR001965; Znf PHD.
 DR InterPro; IPR001941; Znf_ring.
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR SMART; SM00384; AT hook; 1.
 DR SMART; SM00542; FYRC; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00249; PHD; 4.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS0280; SET; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 3.
 DR PROSITE; PS0016; ZF_PHD_2; 3.
 KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
 KW Transcription regulation; Alternative splicing; Repeat.
 FT DNA_BIND 37 44
 FT DNA_BIND 110 117
 FT DNA_BIND 357 365
 FT ZN_FING 959 1005
 FT ZN_FING 1201 1252
 FT ZN_FING 1249 1303
 FT ZN_FING 1335 1396
 FT DOMAIN 1449 1471
 FT DOMAIN 2586 2715
 FT DOMAIN 26 37
 FT DOMAIN 248 255
 FT DOMAIN 362 398
 FT DOMAIN 402 771
 FT DOMAIN 808 812
 FT DOMAIN 1963 1970
 FT DOMAIN 2251 2259
 FT DOMAIN

```
FT VARSPLIC 532 582 VSARSRVKTFRFMDPDKPKVPEVPLRPITPP
FT VQRPAPVPS -> PLSQSLPMTLQSLSLGOWAAPTTS
FT ACLSPLSPULLPRCPLETLGLQ (IN ISOFORM
FT TRUNCATED).
FT MISSING (IN ISOFORM TRUNCATED).
FT K -> E (IN REF. 6).
FT S -> Y (IN REF. 6).
FT E -> Q (IN REF. 6).
FT H -> Y (IN REF. 6).
FT D -> N (IN REF. 6).
FT PLA -> GTR (IN REF. 5; AAH09337).
FT DEE -> ARG (IN REF. 5; AAH07353).
FT D -> H (IN REF. 6).
SQ SEQUENCE 2715 AA; 293511 MW; C0615B981BBEB7BF CRC64;

Alignment Scores:
Pred. No.: 0.0101 Length: 2715
Score: 155.50 Matches: 78
Percent Similarity: 35.64% Conservative: 20
Best Local Similarity: 28.36% Mismatches: 83
Query Match: 10.67% Indels: 94
DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x TRX2_HUMAN (1-2715)
QY 34 CCACTCCGCGCAGCTCCGGGAGAGCGGCACCG-----GCC 72
Db 1810 ProSerGluProGlyGlyGluAspProProLeuAspThrAspValValPro 1829
QY 73 GGCCAGCCCGCCGACCCCATGGA-----CCTCGAGAGGGGACTCGT 114
Db 1830 Gly-AlaProGluArgHisSerProIleGlnAsnLeuAspProProLeuArg----- 1846
QY 115 GGGGATGTTAGCTGCTGCACGGTGTCTGCACCTCCTCGAGTCCAGCTCTCA 174
Db 1847 -----ProAspSerGlySerAlaProProAlaProArgSerPheSe 1861
QY 175 ATCGCACAG-----GGACCCAGGGCGCTGGCCCTCCATCCAGAAA 216
Db 1861 rGlyAlaArgIleValProAsnTyrSerProSerArgProLeuGlyValSe 1881
QY 217 CCTATGACCTCACCGCTACTGGAGCACCAACT-----CCGACGCT 258
Db 1881 rPheGlyProLeuProSerProGlySerProSerSerLeuThrHisIleProThrVa 1901
QY 259 TGGTGGGACCTATCTGAACTACCTGGGCCCCCTTTCAACAGGACGA-----CT 309
Db 1901 lGly-----AspProAspPheProAlaProProArgArgSerArgProSerProle 1919
QY 310 TCAACCTCCCGCTGGGGGCGAGAGCTGCTCCAGGCGCACTGTGACTTGGAGTGT 369
Db 1919 uAlaProArgProProProSerArg-----Tr 1928
QY 370 GGCGAAGCTCAATGACAACTGGGGGTGACCCAGAACTACGAGCCCTACGCCCTTC 429
Db 1928 pAlaSerProLeuLysThrSer-----ProGlnLeuArgValProProThrSe 1946
QY 430 TGTCTTACTTCGTCGCTCAACCTGAGCTGCTGAGCTGCGCCCGCAGCCTGG 489
Db 1946 rValValThrAla-----LeuThrProThrSerGl 1956
QY 490 CCCACTCTGCACCAAGCTCCAGGCGCTGCTGGGCGAGCATTCGGGCGGTCATGGCAGCTC 549
Db 1956 yGluLeu-----AlaProProGlyProAla----- 1964
QY 550 TGGGCTACCCACTGCCCGGCGCTGCTGGGAGTGAACCCACTTGGACTCTGCCCTTG 609
Db 1965 -----ProSerProProPr 1969
QY 610 CCCACAGTACTTCTCCA-----GAAGATGGACGACTTCTGGCTGTGAAGAGCTCC 663
Db 1969 oProGluAspLeuGlyProAspPheGluAspMetGluValValSerGlyLeuSerAlaAl 1989
```

```
QY 664 AGACCTGGCTGTGGCGCTCGGC---CAAGGACTTCAACCGGCTCAAGAGAAGATGCAGC 720
Db 1989 aslpeuacappheAlaAalaserleuLeuGlyThrGluProPheGlnGluLleValAl 2009
QY 721 CTCACAGC-----TGCACTCACCTCCACCTGGGGGCTCA 757
Db 2009 aAlaGlyAlaMetGlySerSerHisGlyGlyProGlyAspSer 2023

RESULT 5
CA44 BOVIN
ID CA44 BOVIN STANDARD; PRT; 453 AA.
AC Q29442;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 4(IV) chain (fragment).
GN COL4A4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
RC TISSUE=Lens;
RX MEDLINE=92112769; PubMed=1370461;
RA Mariyama M., Kalluri R., Hudson B.G., Readers S.T.;
RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
RT collagens."
RL J. Biol. Chem. 267:1253-1258(1992).
RN [2]
RP SEQUENCE OF 217-246.
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen."
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 217-233.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowski R.J., Langeveid J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen."
RL J. Biol. Chem. 262:7874-7877(1987).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DBL: M77480; AAA30458.2; ALT SEQ.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001442; ProcollagenC4.
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD003923; ProcollagenC4; 2.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 FT NON TER 1
 FT DOMAIN <1 222 TRIPLE-HELICAL REGION
 FT DOMAIN 223 453 NON-HELICAL REGION (NC1).
 FT DISULFID 243 332 OR 329 (BY SIMILARITY).
 FT DISULFID 276 329 OR 332 (BY SIMILARITY).
 FT DISULFID 288 294 BY SIMILARITY.
 FT DISULFID 351 449 OR 446 (BY SIMILARITY).
 FT DISULFID 351 449 OR 446 (BY SIMILARITY).
 FT DISULFID 385 446 OR 449 (BY SIMILARITY).
 FT DISULFID 397 404 BY SIMILARITY.
 FT CONFLICT 219 219 I -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 453 AA; 46384 MW; F7ED410AE9A65BC1 CRC64;

Alignment Scores:
 Pred. No.: 0.0118 Length: 453
 Score: 154.50 Matches: 73
 Percent Similarity: 38.08% Conservative: 18
 Best Local Similarity: 30.54% Mismatches: 71
 Query Match: 10.60% Indels: 77
 DB: 1 Gaps: 17

US-09-931-704-1 (1-797) x CA44_BOVIN (1-453)

QY 13 CCGAGCCGCGGCTCGCCCTCCACCTCCGCGAGGAGGAG-----CGCAC 66
 |||:|||||
 DB 18 ProAspProGlyLeu-----ProGlyAspGlnGlyProProGly 30
 |||:|||||
 QY 67 CCGCGCGCGCGAGCCCGCCGCGCGGAGCTCGGAGC-----103
 |||:|||||
 DB 31 ProAspGlyProArgGlyValPro-GlyProProGlyProProGlySerValAspLeu 50
 |||:|||||
 QY 104 -AGGGAGCTCGTGGGGAGTGTAGCTGCT-----GTCACGGTCTCTGGC 150
 |||:|||||
 DB 50 uysGlyGluProGlyAspCysGlyLeuProGlyProProGlyProProGlyProProGly 70
 |||:|||||
 QY 151 ACCTCCC-----TCCAGTCCGAGCTCTCAATCGCACAGGGAGCC 189
 |||:|||||
 DB 70 yProProGlyHisLysGlyPheProGlyCysAsp---GlyLysHisGlyGlnLysGlyPr 89
 |||:|||||
 QY 190 CAGGGCTGCGCCCTCCATCCAGAAACCTATGACTACCCGCTACCTGGAGCACCAC 249
 |||:|||||
 DB 89 oMetGlyPhePro-----GlyProGlnGlyProProGlySerProGly 103
 |||:|||||
 QY 250 TCGCAGCTGGCTGGGACCTATCTGACTTACCTGCGGCGCCCTTTCAACGA-----301
 |||:|||||
 DB 103 yProPro-----GlyAspLysGlyLeuProGlyProProGlyProProGlyProle 120
 |||:|||||
 QY 302 -----GCCAGACTTCAACCTCCCGCTCGGCGGAGAGACTCTCCCGAGGG 348
 |||:|||||
 DB 120 uGlyProProGlySerArgGlyGluProGlyProProAlaAspLeuAspAlaCysProAr 140
 |||:|||||
 QY 349 CCACTTTGAGTGGAGGTGTGGAGAGCTCAATGACAACTCGCGCTGACCCAGAA---406
 |||:|||||
 DB 140 gileProGlyLeuProGlyValProGlyProArg-----GlyProGlu 155
 |||:|||||
 QY 407 -----CTACGAGGCTACAGCACCTTCTGTGTTACTTGTGCTGGCTCAACCGTC 456
 |||:|||||
 DB 155 yThrMetGlyLeuProGlyMetArgGlyPro-----ProGlyPro-- 168
 |||:|||||
 QY 457 AGGCTGCCACTGTGAGCTGCGCGCGAGCCTGGC-----CCTACT 495

Db 169 -GlyCys-----LysGlyGluProGlyLeuAspGlyArgGlyGluAspG 184
 |||:|||||
 QY 496 TGTGCACACCGCTCCAGGCGCTGCTGGCAGCATTTGCGGCGCTCATGGCAGCTCTGGCT 555
 |||:|||||
 DB 184 yLeuProGlySerProGlyProGly---HisLysGlyAspMetGlyGluAlaGlyCy 203
 |||:|||||
 QY 556 ACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCT 608
 |||:|||||
 DB 203 sProGlyAlaProGlyPro-ProGly-----PrometGlyAspProGlyPro 218
 |||:|||||

RESULT 6

CA15 HUMAN
 ID CA15 HUMAN STANDARD; PRT; 1838 AA.
 AC P20908;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(V) chain precursor.
 GN COL5A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.
 RX MEDLINE=91302336; PubMed=2071595;
 RA Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,
 Kato I.;
 RT "Complete primary structure of human collagen alpha 1 (V) chain.";
 RL J. Biol. Chem. 266:13124-13129(1991).
 RN [2]
 RP SEQUENCE OF 621-822.
 RC TISSUE=Chorioamniotic membrane;
 RX MEDLINE=89227189; PubMed=2496661;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of three
 cyanogen bromide-derived peptides from human alpha 1(V) collagen
 chain.";
 RL Arch. Biochem. Biophys. 271:120-129(1989).
 RN [3]
 RP SEQUENCE OF 823-950, AND HEPARIN-BINDING.
 RX MEDLINE=90366601; PubMed=2203476;
 RA Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;
 RT "Primary structure of the heparin-binding site of type V collagen.";
 RL Biochim. Biophys. Acta 1035:139-145(1990).
 RN [4]
 RP SEQUENCE OF 556-571.
 RC TISSUE=Placenta;
 RX MEDLINE=92239022; PubMed=1571108;
 RA Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 characterization by partial sequencing.";
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
 RN [5]
 RP SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.
 RC TISSUE=Chorioamniotic membrane;
 RX MEDLINE=94237164; PubMed=8181482;
 RA Moradi-Ameli M., Rousseau J.C., Klemen J.P., Champliand M.F.,
 Boutilion M.M., Bernillon J., Wallach J.M., van der Rest M.;
 RT "Diversity in the processing events at the N-terminus of type-V
 collagen.";
 RL Eur. J. Biochem. 221:987-995(1994).
 RN [6]
 RP DISEASE, AND VARIANT EDS-I SER-1639.
 RX MEDLINE=97195540; PubMed=9042913;
 RA de Paeppe A., Nuytinck L., Hauser I., Anton-Lamprecht I.,
 Naeyaert J.-M.;
 RT "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos
 syndromes I and II.";
 RL Am. J. Hum. Genet. 60:547-554(1997).
 CC -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE

DR SMART; SM00549; TAPH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 39 42 POLY-HIS.
 FT DOMAIN 52 57 POLY-ALA.
 FT DOMAIN 98 101 POLY-GLY.
 FT DOMAIN 142 148 POLY-ALA.
 FT DOMAIN 268 275 POLY-PRO.
 FT DOMAIN 331 337 POLY-ALA.
 FT DOMAIN 680 683 POLY-PRO.
 FT DOMAIN 808 813 POLY-ALA.
 FT DOMAIN 828 831 POLY-ASP.
 FT CONFLICT 105 117 PGPPSPRRPLVPA -> GRGILLQRRGGRES (IN REF. 3).
 FT CONFLICT 136 136 A -> S (IN REF. 2).
 FT CONFLICT 185 185 G -> GPG (IN REF. 2).
 FT CONFLICT 233 264 MISSING (IN REF. 3).
 FT CONFLICT 293 293 P -> L (IN REF. 3).
 SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Alignment Scores:
 Pred. No.: 0.0126 Length: 1083
 Score: 154.00 Matches: 86
 Percent Similarity: 35.76% Conservative: 17
 Best Local Similarity: 29.86% Mismatches: 108
 Query Match: 10.56% Indels: 77
 DB: 1 Gaps: 16

US-09-931-704-1 (1-797) x T2D3_HUMAN (1-1083)

QY 1 ATTAAGCTTCGCGGAGCGCGG-----CTCGCC 30
 Db 146 ValAlaAlaGlyProGluProAlaProAlaGlyProAlaLysProAlaGlyProAlaAla 165
 QY 31 CTCCACTCCGCCAGCTCCGGAGAGAGCGGCACCGCGCGGCGCCAGCCGCCAGCCCA 90
 Db 166 LeuAlaAlaArgAlaGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 184
 QY 91 TGGACCTCCG-----AGCAGGGGACTCTGGGGGATGTTAGC---GTGCC 132
 Db 185 -GlyLysProAlaGlyProGlyAlaAlaGlnThrLeuAsnGlySerAlaAlaLeuLeuAs 204
 QY 133 TGTGCAGGTGCTTGGCACCCTCCGTCGAGTCCAGTCCAGTCTCAATCGCACAGGGGCCAG 192
 Db 204 nSerHisAlaAlaPro-----AlaValSerLeuValAsnAsnGlyProAl 221
 QY 193 GGCCTGGCCCTCCATCCAGAAACCTATGACCTACCGCTACTTGGAGCCCACTCC 252
 Db 221 aAla-----LeuLeuProLeuProLysProAlaAlaProGly-- 233
 QY 253 GCAGCTTGGCTGGACCTATCTGAACCTACCTGGGCGCCCTTTCAACGAGCCAGACTTCA 312
 Db 234 -----ThrValIleGlnThrProProPheValGlyAlaAlaAlaPr 247
 QY 313 ACCCTCCCG-----CCTGGGCGAGAGACTCTGCCAGGCGCACTGTGTGACT 360
 Db 247 oProAlaProAlaProSerProAlaAlaProAlaProAlaProAlaAlaAla 267
 QY 361 TGGAGGTGTGGGAGCCTCAATGCAACTCGCGCTGACCCAGCACTACGAGGCTTACA 420
 Db 267 aProProProProProProAlaProAlaThrLeuAlaArgProProGlyHisProAlaGl 287
 QY 421 GCCACCTTCTGTGTACTTGGTGGCTTCAACCG-----TCAGGTGGCCACTG 468
 Db 287 yProProThrAlaAlaProAlaValProProProAlaAlaAlaGlnAsnGlyGlySerAl 307
 QY 469 CTGAGTGGCGCGCAGCTTGGCCCACTTCTGCACCGAGCTCCAGGCTCGTGGGAGCA 528
 Db 307 aGlyAlaAlaProAlaProAlaProAlaAlaGlyGlyProAlaGlyValSerGlyGln-- 326
 QY 529 TTGGGGGGTCTGTCAGCTCTGGCTACCCACTGCCCCAGCGCTGCTGGAGCTGAAC 588
 Db 327 ---ProGlyAlaAlaAlaAlaAla----- 335

QY 589 CCACCTTGGACTCTCGCCCTGCCCA-----CA 615
 Db 336 -----ProAlaProGlyValLysAlaGluSerProLysArgValValGl 350
 QY 616 GTGACTTCTCCAGAGATGGACGACTTCTGGCTGTGAGGAGCTGACAGCTGGCT-- 673
 Db 350 nAlaAlaProProAla---AlaGlnThrLeuAlaAlaSerGlyProAlaSerThrAlaAl 369
 QY 674 ----GTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGAGATGCACGCTCCACAG 729
 Db 369 aSerMetValIleGlyProThrMetGlnGlyAlaLeuProSerProAlaAla----- 386
 QY 730 CTCGAGTCACCTCGCACCTGGG 751
 Db 387 -ValProProProAlaProGly 393

RESULT 8
 CA35_HUMAN STANDARD; PRT; 382 AA.
 ID CA35_HUMAN
 AC P25940;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 3(V) chain (fragments).
 GN COL5A3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Placenta;
 EX MEDLINE=92239022; PubMed=1571108;
 RA Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 characterization by partial sequencing."
 RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
 CC -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
 COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
 TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
 CC -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
 MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
 ONE ALPHA 3(V) CHAINS IN PLACENTA.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 1(XI) CHAINS.
 DR PIR; S20375; S20375.
 DR MIM; 120216; -.
 DR InterPro; IPR000087; Collagen.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Collagen.
 FT NON_TER 1
 FT NON_CONS 86 87
 FT NON_CONS 131 132
 FT NON_CONS 167 168
 FT NON_CONS 200 201
 FT NON_CONS 284 285
 FT NON_CONS 317 318
 FT NON_CONS 359 360
 FT NON_TER 382 382
 SQ SEQUENCE 382 AA; 35352 MW; F80BB5968F696B5C CRC64;

Alignment Scores:
 Pred. No.: 0.0169 Length: 382
 Score: 152.00 Matches: 78
 Percent Similarity: 28.84% Conservative: 14
 Best Local Similarity: 24.45% Mismatches: 74
 Query Match: 10.43% Indels: 153
 DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x CA35_HUMAN (1-382)

```
QY 23 GGCTGCGCTCCCTCCCTCCAGCTCCGCGGAGAGAGCCGACCCGCGCGCCAGGCC 82
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16 GlyLeuProGlyHisProGlyLeu***GlyGluGluGlyAlaGlnGlyProGlnGlyPro 35
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 CAG-----CCCATGGACTCGG----- 100
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36 ArgGlyLeuGlnGlyProHisGlyProGlyArgValGlyLysMetGlyArgProGly 55
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 -----AGCAGGGGACTCGTGGGGAT----- 121
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 AlaAspGlyAlaArgGlyLeuProGlyAspThrGlyPro***GlyAspArgGlyPheAsp 75
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 ----- 121
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 GlyLeuProGlyLeuProGlyGlu***GlyGlnGlyProProGlyAsnProGlyIlePro 95
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 -----GTTAGCGTGCTGTGCAC----- 139
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GlyLeuProGlySerAspGlyProLeuGlyHisProGlyHisGlnGlyProThrGlyGlu 115
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 140 ---GGTGCTCTGGCACCTCCCTGCAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCC 196
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 **GlyAlaGlnGlyProProGlySerAla-----GlyProProGly 129
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 197 TGGCCCTCCATCCAGAAAC----- 217
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 TyrProLeuGlnGlyGlu***GlyGlu***GlyGluAspGlyPheProGlyPhe***Gly 149
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 218 -----CTATGACCTCACCGCTACCTGAGGACCACTCCGAGCTTGGCTGG 265
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 AspValGlyLeuLeuGlyAspGlnGlyLysProGlyAlaProGlyProArg---GlyLeu 168
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 266 GACCTATCTGAATCACTCGGCGCCCTCTTCAACGAGCC----- 304
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 GlyValProGlyLeuProGlyTyrProGlyArgProGlyPro***GlySerIleGlyPhe 188
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 ----- 304
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 ProGlyProLeuGlyProIleGlyGluGlyLysThrGlyGluValGlyProLeuGly 208
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 305 ---AGACTTCAACCTCCCGGCTGGG----- 328
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 GluArgGlyProProGlyProProGlyProProGlyGluGlnGlyLeuProGlyLeuGlu 228
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 329 GGCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCTCAATGACAA 388
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 GlyArgGluGlyAla***GlyGlu-----LeuGly----- 238
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 389 ACTGCGGCTGACCCAGAACTACGAGCCCTACAGCCACCTTCTGTGTTACTTGGTGGCT 448
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 -----ProProGlyLeu***GlyLysGluGlyProAlaGlyLeuArgGlyPhePro 255
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 449 CAACCGTACGGTGCACCTGCTGAGTGGCGCGCAGCCCTGGCCCACTT----- 496
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 GlyPro***Gly-----GlyProGlyAspProGlyProThrGlyLeu***Gly 271
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 497 CTGCACACGCTCCAGGGCTGTGTGGGAGCATTTGGGCGTCATCGCAGCTCTGGGCTA 556
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 AspAspGlyProProGlyProVal-----GlyAlaAsnGlySerAspGlyIle 287
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 557 CCCACTGCCCCAGCCCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCC 611
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 ProGly-ProLeuGlyLeuProGly-----ProProGlyAlaAlaGlyProSer 303
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 9
SPD2_NEPCL
ID SPD2_NEPCL STANDARD; PRT; 627 AA.
AC P46804;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spidroin 2 (Dragline silk fibroin 2) (fragment).
OS Nephila clavipes (Orb spider).
```

```
CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
CC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
CC NCBI_taxid=6915;
CC [1]
CC SEQUENCE FROM N.A. PubMed=1527052;
CC MEDLINE=92406876;
CC Hinman M.B., Lewis R.V.;
CC "Isolation of a clone encoding a second dragline silk fibroin.
CC Nephila clavipes dragline silk is a two-protein fiber.";
CC J. Biol. Chem. 267:19320-19324(1992).
CC -!- FUNCTION: Spiders major ampullate silk possesses unique
CC characteristics of strength and elasticity. Fibroin consists of
CC pseudocristalline regions of antiparallel beta-sheet interspersed
CC with elastic amorphous segments.
CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- DOMAIN: Highly repetitive protein characterized by regions of
CC polyalanine and glycine-rich repeating units.
CC -!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M92913; AAA29381.1; -.
CC Silks; Repeat.
CC KW NON TER 1 530 15 APPROXIMATE TANDEM REPEATS.
CC FT DOMAIN 1 36 1.
CC FT REPEAT 1 36 1.
CC FT REPEAT 37 79 2.
CC FT REPEAT 80 121 3.
CC FT REPEAT 122 172 4.
CC FT REPEAT 173 213 5.
CC FT REPEAT 214 252 6.
CC FT REPEAT 253 283 7.
CC FT REPEAT 284 317 8.
CC FT REPEAT 318 359 9.
CC FT REPEAT 360 391 10.
CC FT REPEAT 392 428 11.
CC FT REPEAT 429 464 12.
CC FT REPEAT 465 488 13.
CC FT REPEAT 489 515 14.
CC FT REPEAT 516 530 15.
CC SQ SEQUENCE 627 AA; 54184 MW; CB9B63779B2C594B CRC64;
CC -----
Alignment Scores:
Pred. No.: 0.0168 Length: 627
Score: 152.00 Matches: 79
Percent Similarity: 33.58% Conservative: 12
Best Local Similarity: 29.15% Mismatches: 113
Query Match: 10.51% Indels: 67
DB: 1 Gaps: 13
US-09-931-704-1 (1-797) x SPD2_NEPCL (1-627)
QY 681 AGCGCCACAGCAGGCTCTGACGCTCTTCCAGCCAGAGTCGTCATCTTCTGGAGGA 622
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 SerAlaAlaAlaAlaSerAlaAlaAlaSerAlaGluSerGlyGlnGlnGlyProGlyGly 130
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 621 AGTCACTGTGGCGCGCCAGGAGTCCAGTGGGTTTCAGTCCCGAGCGGCTGGGCA 562
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 TyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGly 150
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 GTGGGT-----AGCCAGAGCTGCCATGACGCCCGCAATGCTGCCAGCA--- 517
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 TyrGlyProGlyGlnGlnGlyProSerGlyProGlySerAlaAlaAlaAlaAla 170
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 516 -----GSCCCTGGA 508
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 171 AlaSerGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGly 190
QY 507 GGCTGG-----TGAGAGTGGGCGGAGCTGGGCGGAGCTCAG----- 469
Db 191 GlyTyrGlyProGlyGlnGlnGlyProSerGlyProGlySerAlaAlaAlaAlaAla 210
QY 468 -----CAGTGGCAGCTCAGCGTTGAGGCGGAGCGCAAGTAACACAGAA 427
Db 211 AlaAlaSerGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyPro 230
QY 426 GGTGGCTGTAGGCT-----CGTAGTCTGGGTGAGCGGAGCTTTGTCTAT 382
Db 231 GlyGlyTyrGlyProGlyGlnGlnGlyLeuSerGlyProGlySerAlaAlaAlaAla 248
QY 381 TGAGGCTTCCGACACCTCCAGTCCAGTGGGCGGAGCTCTGCGGCGGAGC 322
Db 249 ---AlaAlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyTyrGly-----ProGly 265
QY 321 GGGAGGCTTGAAGTCTGGCTCGTTGAAGGGGGGCGGAGCTAGTTCAGATAGTCCGAG 262
Db 266 GlnGlnGlyProSer-----GlyProGlySerAla---AlaAlaAla 278
QY 261 CCAAGCTCGGAGTGTGTCTCCAGTGGGCGGAGTGTAGTCTGAGTGGAGG 202
Db 279 AlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyr 298
QY 201 GGCAGGCCC---CTGGGTCCCTCTGCGATTGAGAGCTGGGAGTGGGAGTCCGAGA 145
Db 299 GlyProGlyGlnGlnGlyPro-----SerGlyAlaGly 309
QY 144 GCACCGTGCAGCAGCAGCTAAACATCCCGGAGTCCCTGCGGAGGTCCTAGGCGGC 85
Db 310 SerAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyLeuGlyGly-TyrGlyPr 329
QY 84 TGGG---GCTGGGCGGCGGCGGCTGGG-----TCCTCTCCGAGGCTGGGCGAG 37
Db 329 oGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyPr 349
QY 36 TGGGAGGCGGAGCGGCGCTCCGCGGAGCT 6
Db 349 oGlySerAlaSerAlaAlaAlaAlaAla 359
RESULT 10
PRP3_MOUSE STANDARD; PRT; 296 AA.
AC P05143;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Proline-rich protein MP-3 (Fragment).
GN PRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86059475; PubMed=2999141;
RA Ann D.K., Carlson D.M.;
RT "The structure and organization of a proline-rich protein gene of a
mouse multigene family."
RL J. Biol. Chem. 260:15863-15872 (1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12100; AAA40005.1; -.
DR MGD; MGI:97773; Prp.

KW Repeat; Saliva.
FT NON TER 1
SQ SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;
Alignment Scores:
Pred. No.: 0.0279 Length: 296
Score: 148.50 Matches: 82
Percent Similarity: 36.92% Conservative: 14
Best Local Similarity: 21.54% Mismatches: 87
Query Match: 10.19% Indels: 77
DB: 1 Gaps: 18
US-09-931-704-1 (1-797) x PRP3_MOUSE (1-296)
QY 12 GCCGAGCGGCGCTGCGCTCCACT-----CGGCGAGCTCCGGG 53
Db 16 SerGlySerGlnProArgProValAenGlySerGlnGlnGlyProPro-ProProGl 35
QY 54 A-----GAGGAGCGGACCCCGCGGC-----CGAGCCCC 83
Db 35 yGlyProGlnProArgProProGlnGlyProProProGlyGlyProGlnProArgPr 55
QY 84 AGCCCATGAGACTCCGACAGGAGGACTCGTGGGGATGTTAGCGTCTGTGACCGGT 143
Db 55 o-ProGlnGlyProProProGly-----GlyProGlnProArgProGlnGlyP 73
QY 144 CTCTGGCAGCTCCCTGCAGTGCAGCTCTCAATCGCACAGGAGGAGCCAGGCGCTGGCCCC 203
Db 73 to---ProProProGlyGlyProGlnProArgProProGlnGlyProProPro---ProG 91
QY 204 TCCATCCAGAAAACCTATGACCTCACCGCTACCTGGAGCCCAACTCCCGAGCTTGGCT 263
Db 91 lyGlyProGln-----ProArgPro---ProGlnGlyProProPro----- 103
QY 264 GGGAGCTATCTGAATACCTGGGCGGCTTTCAACAGAGCAGACTTCAACCTCCCGCG 323
Db 104 -----ProGlyGlyProGlnGlnArgProProGlnGlyProProProp 118
QY 324 CTGGGGG-----AGAGACTCTGCCGAGGCGACTGTTGACTTGGAGGTGGCGAAGC 377
Db 118 roGlyGlyProGlnGlnArgProProGlnGlyProProProProGlyGly----- 134
QY 378 CTCAATGACAAACTCGGCTGACCCAGAACTAGAGGCGCTACAGCCACCTTCTGTGTAC 437
Db 135 -----ProGlnProArgProProGlnGlyPro----- 143
QY 438 TTGCGTGGCTCAACCGTCAAGCTGCGCTGCCACTGTGAGCTGCGCGGAGCTTGGCCACTTC 497
Db 144 -----ProProProAlaGly-----ProGlnProArgPro---P 154
QY 498 TGCACGAGCTCCAGGCGCTGGGCGAGCATGTC-----GGGCGTCATGGCA 545
Db 154 roGlnGlyProProProProAlaGlyProHisLeuArgProThrGlnGlyProProPro 174
QY 546 GGTCTGGG-----CTACCCACTGCGCCCGCGCTGCTCTCTGCTGCTGCT 590
Db 174 hrGlyGlyProGlnGlnArgTyrProGlnSerPro-ProProProGlyGlyProGlnPro 193
QY 591 ACTTGGAGCTCTGGCGCTGCCACAGTACTTCTCCAGAAAGATGGACGACTTCTGGCTG 650
Db 194 ArgProProGlnGlyPro-ProProProGlyGlyProHisProArgProThrGlnGly-- 212
QY 651 CTGAGGAGGCTGCAGACCTGGCTGGGCGCTCGGCGCAAGGACTTCAACCGCG 702
Db 213 -----ProProProThrGlyProGlnProArgProThrGlnGly 225
RESULT 11
CA34 HUMAN STANDARD; PRT; 1670 AA.
ID CA34 HUMAN
AC Q01955; Q9BQ12;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 3 (IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human alpha 3 (IV) collagen chains in
human tissues.";
RL J. Biol. Chem. 269:23013-23017 (1994).
[2]
RN REVISIONS
RP Leinonen A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A., VARIANTS AS E-297; R-407; R-640; R-1167; E-1207;
RP Q-1215; S-1277; T-1330; E-1334; E-1347 AND C-1661, AND VARIANTS R-43;
RP E-162; Y-326; H-408; R-451; L-574; E-1269 AND P-1474.
RX MEDLINE=21064696; PubMed=1134255;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RT Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RL "Structure of the human type IV collagen gene COL4A3 and mutations in
autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106 (2001).
[4]
RN SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=93015826; PubMed=1400291;
RX Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RA "Exon/intron structure of the human alpha 3 (IV) gene encompassing the
Goodpasture antigen (alpha 3 (IV)NCL). Identification of a potentially
antigenic region at the triple helix/NCL domain junction.";
RL J. Biol. Chem. 267:19780-19784 (1992).
[5]
RN SEQUENCE OF 1453-1670 FROM N.A.
RP MEDLINE=91353570; PubMed=1882840;
RX Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;
RA "Sequence and localization of a partial cDNA encoding the human alpha
3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554 (1991).
[6]
RN SEQUENCE OF 1331-1670 FROM N.A.
RP TISSUE=Kidney;
RX Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601 (1992).
[7]
RN SEQUENCE OF 1644-1670 FROM N.A.
RP TISSUE=Kidney;
RX Ding J.;
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
[8]
RN SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RP TISSUE=Kidney;
RX Feng L., Xia Y., Wilson C.B.;
RA "Alternative splicing of the NCL domain of the human alpha 3 (IV)
collagen gene. Differential expression of mRNA transcripts that
predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348 (1994).
[9]
RN SEQUENCE OF 1-29 FROM N.A.
RP MEDLINE=98196854; PubMed=9537506;
RA Monota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3 (IV) and

alpha4 (IV) collagen chains are arranged head-to-head on chromosome
2q36.";
RL FEBS Lett. 424:11-16 (1998).
[10]
RN ALTERNATIVE SPLICING.
RP MEDLINE=93280184; PubMed=8505332;
RX Bernal D., Quinones S., Saus J.;
RA "The human mRNA encoding the Goodpasture antigen is alternatively
spliced.";
RL J. Biol. Chem. 268:12090-12094 (1993).
[11]
RN VARIANT PRO-1474.
RP MEDLINE=95078827; PubMed=7987301;
RX Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Reiders S.T., Smeets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
recessive Alport syndrome.";
RL Hum. Mol. Genet. 3:1269-1273 (1994).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE',
MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
NIDOCEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS. ALPHA 1 (IV) -
ALPHA 6 (IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND
3/15; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
C-TERMINAL NCL DOMAINS.
CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
COCHLEA, LUNG AND BRAIN.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
DOMAIN (NCL) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE
G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
N-LINKED GLYCOSYLATION SITE.
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
THESE, LOCATED IN THE NCL DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
IV COLLAGENS.
CC -1- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
CC -1- DISEASE: ANTIBODIES AGAINST THE NCL DOMAIN OF ALPHA3 (IV) MEDIATE
THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
MALES AND FEMALES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

DR EMBL; X80031; CAA56335.1; --
DR EMBL; AJ288487; CAC36101.1; --
DR EMBL; AJ288488; CAC36101.1; JOINED.
DR EMBL; AJ288489; CAC36101.1; JOINED.
DR EMBL; AJ288490; CAC36101.1; JOINED.
DR EMBL; AJ288491; CAC36101.1; JOINED.
DR EMBL; AJ288492; CAC36101.1; JOINED.
DR EMBL; AJ288493; CAC36101.1; JOINED.
DR EMBL; AJ288494; CAC36101.1; JOINED.

DR EMBL; AJ288495; CAC36101.1; JOINED.
 DR EMBL; AJ288496; CAC36101.1; JOINED.
 DR EMBL; AJ288497; CAC36101.1; JOINED.
 DR EMBL; AJ288498; CAC36101.1; JOINED.
 DR EMBL; AJ288499; CAC36101.1; JOINED.
 DR EMBL; AJ288500; CAC36101.1; JOINED.
 DR EMBL; AJ288501; CAC36101.1; JOINED.
 DR EMBL; AJ288502; CAC36101.1; JOINED.
 DR EMBL; AJ288503; CAC36101.1; JOINED.
 DR EMBL; AJ288504; CAC36101.1; JOINED.
 DR EMBL; AJ288505; CAC36101.1; JOINED.
 DR EMBL; AJ288506; CAC36101.1; JOINED.
 DR EMBL; AJ288507; CAC36101.1; JOINED.
 DR EMBL; AJ288508; CAC36101.1; JOINED.
 DR EMBL; AJ288509; CAC36101.1; JOINED.
 DR EMBL; AJ288510; CAC36101.1; JOINED.
 DR EMBL; AJ288511; CAC36101.1; JOINED.
 DR EMBL; AJ288512; CAC36101.1; JOINED.
 DR EMBL; AJ288513; CAC36101.1; JOINED.
 DR EMBL; AJ288514; CAC36101.1; JOINED.
 DR EMBL; AJ288515; CAC36101.1; JOINED.
 DR EMBL; AJ288516; CAC36101.1; JOINED.
 DR EMBL; AJ288517; CAC36101.1; JOINED.
 DR EMBL; AJ288518; CAC36101.1; JOINED.
 DR EMBL; AJ288519; CAC36101.1; JOINED.
 DR EMBL; AJ288520; CAC36101.1; JOINED.
 DR EMBL; AJ288521; CAC36101.1; JOINED.
 DR EMBL; AJ288522; CAC36101.1; JOINED.
 DR EMBL; AJ288523; CAC36101.1; JOINED.
 DR EMBL; AJ288524; CAC36101.1; JOINED.
 DR EMBL; AJ288525; CAC36101.1; JOINED.
 DR EMBL; AJ288526; CAC36101.1; JOINED.
 DR EMBL; AJ288527; CAC36101.1; JOINED.
 DR EMBL; AJ288528; CAC36101.1; JOINED.
 DR EMBL; AJ288529; CAC36101.1; JOINED.
 DR EMBL; AJ288530; CAC36101.1; JOINED.
 DR EMBL; AJ288531; CAC36101.1; JOINED.
 DR EMBL; AJ288532; CAC36101.1; JOINED.
 DR EMBL; AJ288533; CAC36101.1; JOINED.
 DR EMBL; AJ288534; CAC36101.1; JOINED.
 DR EMBL; AJ288535; CAC36101.1; JOINED.
 DR EMBL; AJ288536; CAC36101.1; JOINED.
 DR EMBL; AJ288537; CAC36101.1; JOINED.
 DR EMBL; AJ288538; CAC36101.1; JOINED.
 DR EMBL; M2993; AAA21610.1; -.
 DR EMBL; M29790; AAA19637.1; -.
 DR EMBL; MB1379; AAA1556.1; -.
 DR EMBL; L08650; AAA52044.1; -.
 DR EMBL; U02519; AAA18942.1; -.

Alignment Scores:

Pred. No.: 0.0275 Length: 1670
 Score: 148.50 Matches: 76
 Percent Similarity: 35.00% Conservative: 15
 Best Local Similarity: 29.23% Mismatches: 72
 Query Match: 10.19% Indels: 97
 DB: 1 Gaps: 18

US-09-931-704-1 (1-797) x CA34_HUMAN (1-1670)

QY 1 ATTAAGCTTCGCGAGCGGCTCGCCCTCCACATC-----CGC 42
 Db 576 VallysglyLeuProglyProlysglyGluLeuAlaLeuSerGlyGlyLeuGlyAspGln 595
 QY 43 CAGCTCCGGGAGAG-----GAGCGCACCGCGCGCGCCAGCC----- 81
 Db 596 GlyProGlyAspProglySerProglySerProAlaGlyProAlaGlyPro 615
 QY 82 CCAGCCCCATGCTCCGACAGG---GGACTCTGGGGGATGTTAGCGTCCCTGTGCA 138
 Db 616 Progly-TyrglyProGlnGlyGluProGlyLeuGlnGlyThrGlnGlyValPro----- 633
 QY 139 CGGTGCTCTGGCACCTCCCTCGATGCGAGCTCTCAATCGCACAGGGGACCCAGGGCGCTG 198

Db 634 -GlyAlaProGlyProGlyGluAla-----GlyProargGly-- 646
 QY 199 GCCCTCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCTCCGAGCT 258
 Db 647 -----GluLeuSerValSerThrProValProGlyProProGlyPro----- 660
 QY 259 TGGCTGGGACCTATCTGAACCTACCTGGGCGCCCTTTCAACGAGCAGACCTTCAACCCCTC 318
 Db 661 -----ProGlyProPro-----GlyHisProGlyProG 670
 QY 319 CCCGCTGGGGCAGAGACTCTGCCAGGCGCAGCTTTGACTTGGAGGTGTGCGGAGCC 378
 Db 670 nGlyProGlyIleProGlySerLeuGlyIleGlySerGlyAspProGlyLeuProGlyPr 690
 QY 379 TCAATGACAAACTGCGGCTGACCCAGACTACGAGCGCTTACAGCCACTCTCTGTGTACT 438
 Db 690 oAsp-----GlyGluProGlyIleProGlyIleGlyPheProGlyPro----- 704
 QY 439 TCGGTGGCTCAACCGTCA-----CAGCTCTGC-----GGCTG 462
 Db 705 -----ProGlyProLysGlyAspGlnGlyPheProGlyThrLysGlySerLeuGlyCy 722
 QY 463 COACTGCTGAGCTGCGCG-----CAGCTCTGC----- 490
 Db 722 s-----ProGlyLysMetGlyGluProGlyLeuProGlyLysProGlyLe 737
 QY 491 -----CCACTTCTGCACCGCTCCAGCGCGCTCT----- 520
 Db 737 uProGlyAlaLysGlyGluProAlaValAlaMetProGlyGlyProGlyThrProGlyPh 757
 QY 521 ----GGGCAG-----CATTGGGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCCGCAGC 570
 Db 757 eProGlyGluArgGlyAsnSerGlyGluHisGlyGluIleGlyLeuProGlyLeuProG 777
 QY 571 CGCTCCCTGGGACTGAA-----CCCATTTGAGCTCTCTGGC 605
 Db 777 y-LeuProGlyThrProGlyAsnGluGlyLeuAspGlyProArgGlyAspProGly 795
 RESULT 12
 ID PRPM_HUMAN STANDARD; PRT; 234 AA.
 AC P10161; P02813;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Salivary proline-rich protein PO (Allele M) [Contains: Peptide P-D] (Fragment).
 GN PRB4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=89121439; PubMed=3220251;
 RA Lyons K.M., Stein J.H., Smithies O.;
 RT "Many protein products from a few loci: assignment of human salivary proline-rich proteins to specific loci."
 RL Genetics 120:255-265 (1988).
 RN [2]
 RP MEDLINE=8186122; PubMed=6041349;
 RA Saitoh E., Isemura S., Sanada K.;
 RT "Complete amino acid sequence of a basic proline-rich peptide, P-D, from human parotid saliva."
 RL J. Biochem. 93:495-502 (1983).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X07704; CAA30542.1; -;
DR PIR; A03295; PIHUSD.
DR PIR; S03175; S03175.
DR MIM; 168730; -;
DR MIM; 180990; -;
KW Repeat; Parotid gland; Saliva; Multigene family.
FT NON TER 1 234 PEPTIDE P-D.
FT CHAIN 165 234
SQ SEQUENCE 234 AA; 23676 MW; 310AFPI3A44E747F CRC64;

Alignment Scores:
Pred. No.: 0.0322 Length: 234
Score: 147.50 Matches: 75
Percent Similarity: 31.62% Conservative: 11
Best Local Similarity: 27.57% Mismatches: 72
Query Match: 10.12% Indels: 114
DB: 1 Gaps: 17

US-09-931-704-1 (1-797) x PRPM_HUMAN (1-234)

Qy 13 CGGAGCGCGGCTGCCCTCCACTCCGCGAGCTCCG----- 51
Db 37 ProProHisProGlyLysProGluArgProProGlnGlyAsnGlnSerGln 56
Qy 52 GGAGGAGCGCCACCGCGCC----- 72
Db 57 GlyProProHisProGlyLysProGluArgProProGlnGlyAsnGlnSer 76
Qy 73 ---GGCCGAGCC-----CAGCCCATGACCTCCGAGCGGAGCTCGTGG 117
Db 77 GlnGlyPro-ProProThrProGlyLysProGluGlyProProGlnGlyAsnGln 96
Qy 118 GATGTTAGCGCTGTGCGAGGTCTG-----GCACCTCCCTGCGAGTCCAGCTC 171
Db 96 nSerGlnGlyProProHisProGlyLysProGluArgProProGlnGlyAsn 116
Qy 172 TCAATCGCAGCGGACCCAGCGCTGGCCCTCCATCCAGAAACCTATGACCTCACCC 231
Db 116 nSerHisArg-----ProProPr 123
Qy 232 GCTACTCGAGACCA-----ACTCCGAGCTTGGTGGAGCTTGAATACCT-- 283
Db 123 oProProGlyLysProGluArgProProGlnGlyAsnGlnSerGlnGlyProPr 143
Qy 284 -----GGGCCCCCTTTCAA-----CGAGCCAGACTTCA 312
Db 143 oProHisProGlyLysProGluGlyProProGlnGlyAsnLysSerArgSerAl 163
Qy 313 ACCCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGTGCG 372
Db 163 aArgSerProProGly----- 168
Qy 373 GAAGCTCAATGACAACTCGGCTGACCAAGACTACAGGCGCTACAGCCACTTCTGT 432
Db 169 -LysProGln-----GlyProGlnGlnGlyAsnLysProGlnGly-- 183
Qy 433 GTTACTGGTGGCTCAACCTGAGGCTGCTGCTGAGTGGCGCCGAGCTGGGCC 492
Db 184 -----ProProProGly-----LysProGlnGlyProPr 194
Qy 493 ACTTCTGACACCGCTTCCAGGCGCTGTGGGAGCATTCGGGCGTCATGCTCTGG 552
Db 194 o-----ProProGly-----GlyAsnProGl 201
Qy 553 GCTACCACTGCCAGCGCTCTCTGAGTGAACCACTTGGACTCTCTGGCCCTGCC 612
Db 201 nGlnProGlnAlaProAlaGly-----LysProGlnGlyProProPr 216
Qy 613 ACAGTGACTTCTCCAGAGATGGAGCACTTCTGGCTGTGAAGAGCTGCAGACTGGC 672

Db 216 o-----ProProGlnGlyArg-----Pr 223
Qy 673 TGTGCGCTCGCCCAAGGACTTCAACCGGCTCAA 706
Db 223 oProArgProAlaGlnGlyGlnProProGln 234
RESULT 13
YQ35 CAEEL
ID YQ35 CAEEL STANDARD; PRT; 317 AA.
AC Q09456;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative cuticle collagen C09G5.5.
GN C09G5.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OK NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Palmer S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z46791; CAA86758.1; -;
DR WormPep; C09G5.5; C01485.
DR InterPro; IPR002486; Col.cuticle_N.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col.cuticle_N; 1.
KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
KW Multigene family; Collagen.
FT DOMAIN 92 124 TRIPLE-HELICAL REGION.
FT DOMAIN 137 199 TRIPLE-HELICAL REGION.
FT DOMAIN 202 264 TRIPLE-HELICAL REGION.
SQ SEQUENCE 317 AA; 31283 MW; 685DCF24612707BB CRC64;

Alignment Scores:
Pred. No.: 0.0321 Length: 317
Score: 147.50 Matches: 68
Percent Similarity: 36.97% Conservative: 10
Best Local Similarity: 32.23% Mismatches: 95
Query Match: 10.12% Indels: 38
DB: 1 Gaps: 10

US-09-931-704-1 (1-797) x YQ35 CAEEL (1-317)

Qy 24 GCTCGCTCCCTCCACTCCGCGCA---GCCTCCGGGAGAGGAGCGCACCCGCGGCCAGC 80
Db 91 AlaGlyProProGlyProGlyAlaSerGlyAspArgGlyLeuAspGlyGlnProGly 110
Qy 81 CCCAGCCCCATGAGCTCCGAGCAGGAGACTCGTGGGGATGTTAGCTGCTGTGCAG 140
Db 111 ProAlaGlyLys--ProGlyGlnProGlyValAlaGly-----ProAlaHis 126
Qy 141 GTGCTCTGGCACCTCCCTGCTGCTGCTCTCTCAATCGCAGGAGGAGCCGCGCTGCG 200

Db 126 isGln---GlnGlnGluCysIleLysCysProGlnGlyAlaProGlyProAlaGlyAlaP 145
QY 201 CCTCCATCCAGAAAACCTATGACTCACCCTACCTGGAGCACCACCACTCGCAGCTTG 260
Db 145 roGlyAsnProGlyProGlnGlyProGlnGlyAsnProGlyAlaProAlaHisGlyGly 165
QY 261 GCTGGGACCTATCTGAATCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCC 320
Db 165 ly-----GlnGlyPro-----ProGlyP 171
QY 321 CGCTGGGGGAGAGACTCTCCAGGCGCCACTGTTGACTGGAGGTGGGAAGCCTC 380
Db 171 roProGlyProAlaGlyAspAlaGlySerProGlyGlnAlaGlyAlaProGlyAsnProG 191
QY 381 AATGACAACTCGCGCTGACCCAGAACTACGAGGCTCAGCCACTCTCTGTGTTACTTG 440
Db 191 lyArgProGlyGlnSerGlyGlnArgSerArgGlyLeuProGlyProSer-----G 208
QY 441 CGTGGCTCAACCTCAGGCTGCCACTGCTGAGCTGCGCGC----- 481
Db 208 lyArgProGlyProGlnGlyProPro-----GlyAlaProGlyGlnProGlySerGlyS 226
QY 482 --CAGCTGGCGCCACTT---CTGACACGAGCTCCAGGCGCTGCTGGCAG-----CA 530
Db 226 erThrProGlyProAlaGlyProProGlyProGlyProGlyProAsnGlyGlnProGlyHisP 246
QY 531 GCGGGCGCTCATGGCAGCTCTGGGCTACCCACTGCTGCCAGCCGCTGCCCTGGGACTGAACCC 590
Db 246 roGlyGlnAspGlyGlnProGlyAlaProGlyAsnAspGlyAla-ProGlySerAspAla 265
QY 591 ACTTGGACTCTGGCGCTGCCACAGT 617
Db 266 AlaTyrCysProCysProAlaArgSer 274

RESULT 14
ID GPI_CHLRE STANDARD; PRT; 555 AA.
AC Q9FPQ6; Q03927;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
DE Glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -!- FUNCTION: Major component of the outer cell wall w6 (crystalline)
CC layer.
CC -!- SUBUNIT: Associates with GP2 and GP3.
CC -!- PTM: N-glycosylated and O-glycosylated.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF309494; AAG45420.1; -;
DR EMBL; M58496; AAG49706.1; ALT_SEQ.
DR GlycoSuiteDB; Q9FPQ6; -;
DR InterPro; IPR002965; P rich extensin.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSILEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29
FT CHAIN 30 555
FT DOMAIN 40 339
FT DOMAIN 259 279
FT CARBOHYD 399 399
FT CARBOHYD 455 455
FT CARBOHYD 493 493
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;
Alignment Scores:
Pred. No.: 0.032 Length: 555
Score: 147.50 Matches: 65
Percent Similarity: 33.04% Conservative: 9
Best Local Similarity: 29.02% Mismatches: 103
Query Match: 10.12% Indels: 47
DB: 1 Gaps: 7
US-09-931-704-1 (1-797) x GPI_CHLRE (1-555)
QY 1 ATTAAGCTTCGCGGAGCGCGCTCCCTCCACTCGCCAGCTCCGGGAGAGGAG 60
Db 167 ValProSerProAlaProSerProThrProSerProSerProValPro 186
QY 61 CCGACCGCGCGCGCCAGCCCGATGAGCTCCGAGCGGGGACTCGTGGGGGA 120
Db 187 ProSerProAlaProSerProAlaPro-ProValProSerProAlaProProSe 206
QY 121 TGTAGCGTGTGTCACGGTCTGTCACCTCCCTGCGAGTCCAGC----- 169
Db 206 rProAlaProValProSerProAlaProSerProSerProAlaProPr 226
QY 170 -----TCTCAATGCGACAGGGGACCCAGGGCC----- 196
Db 226 oSerProProAlaProSerProSerProProAlaProSerProValPr 246
QY 197 ----TGGCCCTCCATCCAGAAAACCTAGCTACCCGCTACCTGGAGCACCCTCC 252
Db 246 oProSerProAlaProSerProAlaProSerProSerProLysProAlaProPr 266
QY 253 GCAGCTTGGCTGGGACCTATCTGAACCTAGTGGGCCCCCTTT---CAACGAGCCAGACT 309
Db 266 oProSerProProProProProProProProProProPheProAlaAsnThrProMe 286
QY 310 TCACACCTCCCGCTGGGGGAGAGACTTGTCCCGAGGCCACTGTTGACTTGGAGGTGT 369
Db 286 tProSerProProSerProProProPro----- 297
QY 370 GCGGAGGCTCAATGACAACTGCGGCTGACCCAGCACTACGAGGCTACAGCCACCTTC 429
Db 298 -AlaProThrProThrProThrProSerProSerProSerProValProProSe 317
QY 430 TGTGTACTTGGTGCGCTCAACCGTCAGGCTGCCACTGTGAGCTGCGCGGAGCTGG 489
Db 317 rProAlaProValProSerPro-----AlaProProSerProAl 331
QY 490 CCACCTTGTGACACGCTCCAGGCGCTGCTGGGAGCATTTGCGGGCTCATGGCAGCTC 549
Db 331 aPro-----SerProProSerProAlaPro----- 340
QY 550 TGGGCTACCCACTGCCCCAGCGCTGCTGGGAGTGAACCCACTTGGACTCTGGCCCTG 609
Db 341 -----ProThr-ProSerProSerProSerProSerProSerProSerProS 358

QY 610 CCCACAGT 617
 Db 358 erProSer 360

RESULT 15
 DRPL RAT STANDARD; PRT; 1183 AA.

AC P54258;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Atrophin-1 (Dentatorubral-pallidolysian atrophy protein).
 GN DRPLA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cerebellum, and Striatum;
 EX MEDLINE=97317138; PubMed=9173396;
 RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
 RA Ashworth R.G., Ross C.A.;
 RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
 RT homologue";
 RL Neurobiol. Dis. 2:129-138(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
 RX MEDLINE=96081227; PubMed=8541849;
 RA Schmitt I., Epplen J.T., Riess O.;
 RT "Predominant neuronal expression of the gene responsible for
 RT dentatorubral-pallidolysian atrophy (DRPLA) in rat";
 RL Hum. Mol. Genet. 4:1619-1624(1995).
 CC -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
 CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
 CC -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
 CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U31777; AAA80337.1; -;
 DR EMBL; X89453; CAA61623.1; -;
 DR InterPro; IPR002951; Atrophin.
 DR Pfam; PF03154; Atrophin-1; 2.
 DR PRINTS; PR01222; ATROPHIN.
 FT DOMAIN 165 171 POLY-PRO.
 FT DOMAIN 303 306 POLY-PRO.
 FT DOMAIN 377 383 POLY-SER.
 FT DOMAIN 387 391 POLY-SER.
 FT DOMAIN 440 446 POLY-PRO.
 FT DOMAIN 477 480 POLY-HIS.
 FT DOMAIN 481 489 POLY-GLN.
 FT DOMAIN 502 505 POLY-PRO.
 FT DOMAIN 562 572 POLY-SER.
 FT DOMAIN 702 705 POLY-PRO.
 FT CONFLICT 455 455 N -> S (IN REF. 2).
 FT CONFLICT 594 594 F -> L (IN REF. 2).
 FT CONFLICT 689 689 P -> R (IN REF. 2).
 FT CONFLICT 717 717 T -> M (IN REF. 2).
 FT CONFLICT 737 737 A -> V (IN REF. 2).
 FT CONFLICT 965 965 MISSING (IN REF. 2).
 SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Alignment Scores:
 Pred. No.: 0.0341 Length: 1183

Score: 147.00 Matches: 78
 Percent Similarity: 36.29% Conservative: 12
 Best Local Similarity: 31.45% Mismatches: 66
 Query Match: 10.08% Indels: 93
 DB: 1 Gaps: 13

US-09-931-704-1 (1-797) x DRPL_RAT (1-1183)

QY 39 CCGCCAGCTCCGGAGAGAGAGCCGCCGCC----- 72
 Db 502 ProPro-ProProGlyAlaTyrProHisProLeuGluSerSerAsnSerHisAlaHi 521
 QY 73 -----GCCCCAGCCGCCGCCGCC 89
 Db 521 sProTyrAsnMetSerProSerLeuGlySerLeuArgProTyrProProGlyProAla-H 541
 QY 90 ATGGACCTCCGAGAGAGAGAGCTCTGGGGGATTTAGCGTCTGTGACGGTGTCTCTGG 149
 Db 541 isLeuProProSerHis-----GlyGlnValSerTyrSerGlnAlaGlyProAsnG 558
 QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCA----- 180
 Db 558 lyProProValSerSerSerSerSerSerGlySerSerGlnAlaAlaTyrSerC 578
 QY 181 -----CAGGGGCCGCCGCCGCCGCC-----CCTCCATCCA 211
 Db 578 ysSerHisProSerSerSerGlnGlyProGlnGlyAlaSerTyrProPheProValP 598
 QY 212 GAAACCTATGACCTCACCCTACCTACCTGAGCACC-----AACTCCG 253
 Db 598 roProileThrThrSerSerAlaThrLeuSerThrValleAlaThrValAlaSerSerP 618
 QY 254 CAGCTTGGCTGGGACCTATCTGAACCTACTCTGGGGCCCCCTTTCAACGAGCCAGACTTCAA 313
 Db 618 roAla-GlyTyrLysThrAlaSerProProGlyProPro-----GlnTyrSerLysArgAla 636
 QY 314 CCTCC-----CCGCTGGGGGAGAGACTCTGCCCGAGGC 349
 Db 637 ProSerProGlySerTyrLysThrAlaThrProProGlyTyrLysProGly----- 653
 QY 350 CACTGTTGACTGGAGGTGTGCGAGGCTCAATGACAACTCGCGCTACCCAGCACTA 409
 Db 654 -----SerProProSerPheArgThrGlyThrProGlyTyr 666
 QY 410 CGAGGCTACAGCCACCTTCTGTGTTACTTGGTGGCTCAACCGTCAGGCTGCCACTGC 469
 Db 667 ArgGlyThrSerProPro-----AlaGlyPro-----GlyThrPheLys 679
 QY 470 TGAGTGGCGCGCAG-----CCTGGCCCACTTCTGACACGCTCCAGGCGTGTCTGGG 523
 Db 680 ProGlySerProThrValGlyProGlyProLeu-----ProProAlaGlyProSer--- 696
 QY 524 CAGCATTGGCGGCGTCATGGGAGCTCTGGGCTACCCACTGCCCGCCAGCCGCTGCTGGAC 583
 Db 697 -----SerLeuSerSerLeuProProProProAlaAla-ProThrTh 710
 QY 584 TGAACCCACTTGGACT 599
 Db 710 rglyProProLeuThr 715

Search completed: January 27, 2003, 16:02:37
 Job time : 22.3445 secs

2

RP SEQUENCE FROM N.A.
RX MEDLINE=9382254; PubMed=10448081;
RA Shi Y., Wang W., Youre P.A., Gohari S., Zukauskas D., Zhang J.,
R Ruben S., Alderson R.F.,
RT "Computational EST database analysis identifies a novel member of the
RL neuropeptide cytokine family,"
RN Biochem. Biophys. Res. Commun. 262:132-138 (1999).
[3]
RP SEQUENCE FROM N.A.
RA Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA TISSUE=KIDNEY;
RC Scrausberg R.,
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176912; AAF00992.1; -;
DR EMBL; AF172854; AAD54284.1; -;
DR EMBL; AF176911; AAF00991.1; -;
DR EMBL; AY049779; AAL15436.1; -;
DR EMBL; BC012939; AAL12939.1; -;
SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Alignment Scores:
Pred. No.: 4.72e-87 Length: 225
Score: 1226.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 4 Gaps: 0

US-09-931-704-1 (1-797) x Q9UBD9 (1-225)
QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTGCTGCTGCACGGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACTCCCTGAGTGCAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCCTGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40
QY 210 CAGAAACCTATGACCTACCGCTACCTGAGGACCACTCCGACGAGTGTGGCTGGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGAGCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 330 GCAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnArg 100
QY 390 CTGGGCTGAGGCTGACCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCTGAGGCTGACCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGCTGTGGGAGCAGTTCGGGCGCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
QY 570 CCGCTGCTGGGAGTGAACCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGACGAGTCTTGGCTGCTGAAGGAGTGTGAGAGCTGGGCTGGGCGCTGGGCGCAAG 689
Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 690 GACTTCACCGCTCAAGAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGACCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GlyAlaHisGlyPhe 225

RESULT 2
Q9QZM3 PRELIMINARY; PRT; 225 AA.
ID Q9QZM3;
AC Q9QZM3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Neurotrophin-1/B-cell stimulating factor-3.
GN BSF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Szamien U., Starnes C., Lile J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.,
RT "Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
RT IL-6 family,"
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463 (1999).
DR EMBL; AF176913; AAF00993.1; -;
DR MGD; MGI:1930988; Bsf3.
SQ SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;

Alignment Scores:
Pred. No.: 1.72e-84 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 81.82% Indels: 0
DB: 11 Gaps: 0

US-09-931-704-1 (1-797) x Q9QZM3 (1-225)
QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTGCTGTCACGGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACTCCCTGAGTGCAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCCTGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40
QY 210 CAGAAACCTATGACCTACCGCTACCTGAGGACCACTCCGACGAGCTTGGG 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGAGCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 330 GCAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnArg 100
QY 390 CTGGGCTGAGGCTGACCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCTGAGGCTGACCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGCTGTGGGAGCAGTTCGGGCGCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
QY 570 CCGCTGCTGGGAGTGAACCTGAGTGGCGCGCAGCTTCAACCTCCGACGAGCTTGGG 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGACGAGTCTTGGCTGCTGAAGGAGTGTGAGAGCTGGGCTGGGCGCTGGGCGCAAG 689
Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

```

Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaThrLeuGlyTyProLeuProGln 160
QY 570 CGCTGCTGGGAGTGAACCCACTTGGACTCTGGCCCTGCCACAGTACTCTCCAG 629
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGGAGGACTTCTGGCTGCTGAAGAGCTGCAGACCTGGCTGGGCTCGGCCAAG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGCTCAAGAAGATGACAGCTCCAGCAGCTGCAGTACCTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GluAlaHisGlyPhe 225

RESULT 3
Q9QZR9 PRELIMINARY; PRT; 1682 AA.
AC Q9QZR9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Alpha 4 collagen IV.
GN COL4A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
RT mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169388; AAD50450.1; -.
DR MGD; MGI:104687; Col4a4.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001442; ProcollagenC4.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 22.
DR ProDom; PD000007; Collagen; 4.
DR ProDom; PD003923; ProcollagenC4; 1.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1682 AA; 164096 MW; 6F7B679EDD76E904 CRC64;

Alignment Scores:
Pred. No.: 1.66e-05 Length: 1682
Score: 177.50 Matches: 77
Percent Similarity: 37.60% Conservative: 14
Best Local Similarity: 31.82% Mismatches: 81
Query Match: 12.17% Indels: 70
DB: 11 Gaps: 13

US-09-931-704-1 (1-797) x Q9QZR9 (1-1682)

QY 12 GCGGAGCGCGCTCGCCCT-----CCCACTCGGCAGCCTCGGGAGA 56
Db 1235 AlaGlyAlaProGlyArgAlaAlaLysGlyAspIleProAspPro-GlyProProGlyAs 1254
QY 57 GGAG-----CCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103
Db 1254 pArgGlyProProGlyProAspGlyProArgGlyValPro-GlyProProGlySerProG 1274
QY 104 -----AGGGGACTCGTGGGGATGTTAGCGTGCCT----- 133

```

```

Db 1274 lYasnValAspLeuLeuLysGlyAspProGlyAspCysGlyLeuProGlyProGly 1294
QY 134 -----GTGCACGGTGTCTTGGCACCTCCCTGTCAGTGCAGCTCTCAAT 176
Db 1294 eRArgGlyProProGlyProGlyCysGlnGlyProGlyProGlyCysAspGlyLysAspG 1314
QY 177 GCGACAGGGAGCCAGGGCCCTGCGCCCTCCATCCAGAAAACCATATGACCTACCCGCTAC 236
Db 1314 lYglnLysGlyProMetGlyLeuProGlyLeuProGly-----ProProGlyLeuP 1331
QY 237 CTGGAGGACCAACTCCGAGCTTGGCTGGGACCTATCTCAACTACTGGGCGCCCTTTC 296
Db 1331 roGlyAlaProGlyLys-----GlyLeuProGlyProProGlyA 1345
QY 297 AACGA-----GCCAGACTTCAACCCCTCCCGCTCGGGGCGAGAG 335
Db 1345 lGlyGlyProValGlyProProGlyCysArgGlyGluProProAlaAspValA 1365
QY 336 ACTTGTCCCGAGGCGCACTTGTGACTTGGAGTGTGGCGAAGCCTCAATGACAACTGCGG 395
Db 1365 spSerCysProArgIleProGlyLeuProGlyValProGlyProArgGlyProGluGlyA 1385
QY 396 CT-----GACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGTGGCCTC 449
Db 1385 lAMetGlyGluProGlyArgArgGlyLeuProGlyPro----- 1397
QY 450 AACCGTCAAGGCTGCACTGCTGAGCTGCGCGCGAGCTGCGCA----- 493
Db 1398 -----GlyCys-----LysGlyGluProGlyProAspGlyArgArgGlyG 1411
QY 494 -----CTTGTGACCAAGCCTCCAGGCGCTGCTGGGCGAGCATTCGCGCGCTCATGGCAGCT 548
Db 1411 lNAspGlyIleProGlySerProGlyProGlyProGlyArgLys---GlyAspThrGlyGluA 1430
QY 549 CTGGGCTACCACTGCGCGCGCGCTGCTGGGAGTGAACCCACTTGGACTCTGGGCGCTC 608
Db 1430 laGlyCysProGlyAlaProGlyPro-ProGly-----ProThrGlyAspProGlyPro 1447

RESULT 4
Q95JDO PRELIMINARY; PRT; 511 AA.
AC Q95JDO;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Basic proline-rich protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]_TaxID=9823;
RP SEQUENCE FROM N.A.
RC TISSUE=PAROTID GLAND;
RA Zhang Q., Szalay A.A., Kyeyune-Nyombi E., Sands J.F., Oberg K.C.,
RA Tieche J.-M., Leonora J.;
RT "Cloning and expression of a novel proline-rich protein from porcine
RT parotid glands.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY035848; AAK61382.1; -.
DR PRINTS; PRO1582; KV33CHANNEL.
SQ SEQUENCE 511 AA; 48483 MW; AB04597964C448D7 CRC64;

Alignment Scores:
Pred. No.: 3.27e-05 Length: 511
Score: 172.50 Matches: 72
Percent Similarity: 37.31% Conservative: 3
Best Local Similarity: 35.82% Mismatches: 79
Query Match: 11.83% Indels: 47
DB: 6 Gaps: 9

US-09-931-704-1 (1-797) x Q95JDO (1-511)

QY 30 CCTCCCACTCCGCGACCTCCGGGAGAGGAGCGGCAC-----CCGCGCGCGCCAGC 80

```


Query Match: 11.48% Indels: 77
DB: 5 Gaps: 16
US-09-931-704-1 (1-797) x Q9BIU6 (1-444)
QY 750 CCAGTGCAGGTGACTGACGCTGCTGAGGCTGCATCTTCTTGTAGCCGGTTGAAGT 691
Db 19 ProGlySerGly- - - - - GlyGlnGlnGlyProGlyGlyGlnGly 31
QY 690 CCTTGG- - - - - CCGAGCGCCACAGCAGGTCTGCAGCT- - - - - CTTCA 652
Db 32 ProGlyGlyProSerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyTyrglyProGly 51
QY 651 GCAGCAGGAAGTCTCTTCTTGGAGAGTCACTGTGGCAGGCGCCAGGAGTCCAG 592
Db 52 AlaGlyGlnGlnGlyProGlySerGlyGlyGlnGlnGlyGlyGlnGlnGln 71
QY 591 TGG- - - - - 589
Db 72 GlyProGlyGlyAlaGlyGlnGlyGlyProArgGlyGlnGlyProTyrglyProGlyAla 91
QY 588 GTTCAGTCCAGGCGAGCTGGGCGAGTGGGTAGCCAGAGCTGCCATGACGCCGCAA 529
Db 92 AlaAlaAlaAlaAlaAlaAlaGlyTyrgly- - - - - ProGly 104
QY 528 TGCTGCCAGCAGGCGCTGGAGGCTGTGTGAGAAGTGGGCCAGGCTGGCGCGAGCTCAG 469
Db 105 AlaGlyGlnGlnGlyProGlySer- - - - - GlnGlyProGlySerGlyGlyGlnGln 121
QY 468 CAGTGGCAGCTGACGGTTGAGGCGCACGCAAGTAAACAGAGGTGGTGTAGGCTGT 409
Db 122 - - - - - GlyProGlySerGlnGlyProTyrglyProSer 132
QY 408 AGTCTGTGGTACGCGCAGTTGTTCATTGAGGCTTCGCCACACCTCCAGTCAACAGTGG 349
Db 133 AlaAlaAlaAlaAlaAlaAlaGlyProGlyTyrglyProGlyAlaGlyGlnGlnGly 152
QY 348 CCTCGGCAGAG- - - - - TCTCTGCCCGGCGGAGGCT- - - - - 313
Db 153 ProGlySerGlnGlyProGlySerGlyGlyGlnGlnGlyProGlyGlyGlnGlyProTyrgly 172
QY 312 - - - - - TGAAGTCTGGCTGTTGAAGGGGGGCGCCAGGTAGTTCAGATAGTCCCG 262
Db 173 GlyProSerAspAlaAlaAlaAlaAlaAlaAlaGlyProGlyTyrgly- - - - - 188
QY 261 CCAAGCTCGGAGTGGTCTCCAGTACGCGGTGAGGTCAATAGTTTCTGATGGAGG 202
Db 189 ProGlyAlaGlyGlnGlnGlyProGlySerGly- - - - - GlyGlnGlnGlyGlyGlnGlySer 207
QY 201 GCGCAGGCGCTGGTCCCTGTGCGATTGAGAGCTGGCAGTCCAGGAGGTGCCAGAGCA 142
Db 208 GlyGlnGln- - - - - GlyPro- - - - - GlyGlyAlaGlyGlnGlyGlyProArgGly 222
QY 141 CCGTGACA- - - - - GGCAGCGTAAATCCCCAGTCCCTGCTGAGAGTCCATGGGCG 85
Db 223 GlnGlyProTyrglyProGlyAlaAlaAlaAlaAlaAlaAlaAlaGlyGly- - - - - TyrglyPr 242
QY 84 TGGGCTGGGCGCGCGGTGGGCTCC- - - - - TCTCCCGA- - - - - GGTGGCGGAGTGGGAG 31
Db 242 oGlyAlaGlyGlnGlnGlyProGlySerGlnGlnGlyProGlySerGlyGlyGlnGlnGlyPr 262
QY 30 GCGCAGCGCGCG 18
Db 262 oGlySerGlnGly 266
RESULT 10
Q02802 ID Q02802 PRELIMINARY; PRT; 1532 AA.
AC Q02802;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Autoantigen (Fragment).

GN BP180.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FORESKIN;
RX MEDLINE=92381323; PubMed=1324962;
RA Giudice G.J., Emery D.J., Diaz L.A.;
RT "Cloning and primary structural analysis of the Bullous pemphigoid
RL autoantigen, BP180.", 99:243-250 (1992).
RL J. Invest. Dermatol.
DR EMBL; M91669; AAA35605.1; -;
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 8.
FT NON TER 1
SQ SEQUENCE 1532 AA; 154568 MW; BF35054CF93BBEE3 CRC64;
Alignment Scores:
Pred. No.: 0.000153 Length: 1532
Score: 165.00 Matches: 77
Percent Similarity: 29.41% Conservative: 23
Best Local Similarity: 22.65% Mismatches: 96
Query Match: 11.32% Indels: 144
DB: 4 Gaps: 13
US-09-931-704-1 (1-797) x Q02802 (1-1532)
QY 20 CCGCGCTCGCCCTCCACTCCGCCAGCTCCGGGAGAGGAGCGGACCCGCGGCCGAG 79
Db 667 ArgGlyGluAlaGlyProProGlySerGlyGlyGlyGlyGlyGlyAlaAlaGlyGlu 686
QY 80 CCGCAGCCCCATGGACCTCCGAGC- - - - - 103
Db 687 ProGlyProHisGlyProProGlyValProGlySerValGlyProGlySerSerGly 706
QY 104 - - - - - AGGGGACTCTGTGGGGATGTT 124
Db 707 SerProGlyProGlnGlyProProGlyProValGlyLeuGlnGlyLeuArgGlyGluVal 726
QY 125 AGCTGCGCT- - - - - GTGCACGGTGTCTGTGGACCTCCCTGTCAGTGCACCTCT 172
Db 727 GlyLeuProGlyValGlySerGlyAspGlyProMetGlyProProGlyProGlyProGlyAsp 746
QY 173 CAATCGCAGCGGAGCCAGGCGCTGGCCCTCCATCCAGAAAACCTATGACTCACCCG 232
Db 747 GlnGlyGlnGlySerGlyProArgGlyLeuThrGlyGluProGly- - - - - MetArgGly 763
QY 233 CTACCTGAGGACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCCC 292
Db 764 LeuProGlyAla- - - - - ValGlyGluPro 771
QY 293 TTTCAACAGCGCAGACTTCAACCTCCCGCTGGGGGAGAGAGACTCTCCAGGGGCCAC 352
Db 772 GlyAlaGlySerGlyAlaMetGlyProAlaGlyProAspGlyHisGlnGlyProArgGlyGlu 791
QY 353 TGTTGACTTGGAGGTGGCGAAGCTCAATGACAACTGCGGCTGACCCAGAACTACGA 412
Db 792 GlnGlyLeuThrGlyMet- - - - - ProGlyIleArg 801
QY 413 GGCCTACAGCCACCTTCTGTGTACTTGTGGCTTCAACCGCTCAGGCTGCCACTGCTGA 472
Db 802 Gly- - - - - ProGly- - - - - ProSerGly- - - - - 808
QY 473 GCTGCGCGCAGCTGGC- - - - - CCATTCTGACACGCTCCAGGCGCTCTCTGGG- - - - - 523
Db 809 - - - - - AspProGlyLysProGlyLeuThrGlyProGlnGlyProGlnGlyLeuPro 825
QY 524 CAGCATTGGGGGTCATGGCAGCTCTGGCTACCACTGCCCCAGCCGCTGCCTGGGAC 583
Db 826 GlyThrProGlyArgProGlyIleGlyGlyGlyGlyGlyGlyAlaPro- - - - - 840

Best Local Similarity: 30.96% Mismatches: 78
Query Match: 11.27% Indels: 71
DB: 5 Gaps: 14

US-09-931-704-1 (1-797) x Q9BIT5 (1-563)

```
QY 684 CCAGAGCCAGCAGCGTCTGCAGCTCTCCACGAGCCAGAGTCGTCCTCTGGA 625
DB 255 ProGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyProGly 274
QY 624 GGAAGTCACTGTGGGAG-----GGCAGAGTCCAAAGTGGTTCAGTCCAGGCA 574
DB 275 GlyTyrGlyProGlyGlnGlnGlyProGlyGlnGlnGlyProSerGlyProGlySerAla 294
QY 573 GCGGCTGGGAGTGGGTAGCCAGAGCTGCCATGACGCCCGCATGTGTCGCCAGCAGC 514
DB 295 AlaAlaAlaAlaAlaAlaGlyProGly-----ProGlnGly 306
QY 513 CTTGGAGCTGGTGCAGAGTGGGCGGAGCTGCGGCGCAGCTCAGCAGTGGCAGCTGAC 454
DB 307 ProGlyGlyTyr-----GlyProGly-----GlnGlnGlyPro--- 317
QY 453 GGTGAGCCAGCAGCAAGTAACACAGAGTGGCTGTAGCTGTGAGTTCGTGGTCAACC 394
DB 318 -----GlyGlyTyrGlyProSerGlyProGlySerAla 328
QY 393 GCAGTTGTGTCATTGAGGCTTCCGACACCTCCAAAGTCAACAGTGCAGCTGGGCGAGCT 334
DB 329 AlaAlaAla-----AlaAlaAlaGlyProGlyGlnGln----- 340
QY 333 CTGCCCCAGCGGGAGGGTTGAAGTCTGGCTGTGTTGAAAGGGG-----GGCCCCAGGTAGT 277
DB 341 --GlyProGlyGlyTyrGlyProGlyGlnGlnArgProSerGlyTyrProGlyGln 359
QY 276 TCAGATAGTCCAGCAGCAGCTGGAGTGGTGGTCCAGTGGGAGTGGTCAATAGG 217
DB 360 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaGlyProGlyGlnGln 379
QY 216 TTTTCTGATGGAGGGGCGAGCGCTGGGTCCTGCTGCGATTGAGAGTGGCACTGCAG 157
DB 380 GlyProGlyAlaTyrGlyProSer---GlyPro----- 389
QY 156 GGAGTCCAGAGCAGCGTGCAGAGCAGCTAACATCCCCACAGTCCCTGTCTCGGA 97
DB 390 --GlySerAlaAlaAlaAlaAlaGly-----LeuGly 399
QY 96 GGTCCATGGG---GCTGGGGTGGGCGCGCGGGTGGCGCTCCTCT----- 54
DB 400 Gly-TyrGlyProAlaGlnGlnGlyProSerGlyAlaGlySerAlaAlaAlaAlaAla 419
QY 53 -----CCCGAGGCTGG-----CGGAGTGGGAGGGGAGCGCGGCTCC 15
DB 419 AlaGlyProGlyGlyTyrGlyProValGlnGlnGlyProSerGlyProGlySer 437
```

RESULT 14

```
O22514 PRELIMINARY; PRT; 326 AA.
ID O22514;
AC O22514;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Proline rich protein.
OS Santalum album (white sandalwood).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Santalales; Santalaceae; Santalum.
OX NCBI_TaxID=35974;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya A., Sita L.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF020261; AAB70928.1;
SQ SEQUENCE 326 AA; 33407 MW; B10011C45A86DCD7 CRC64;
```

Alignment Scores:

Pred. No.: 0.000216 Length: 326
Score: 161.50 Matches: 63
Percent Similarity: 36.87% Conservative: 17
Best Local Similarity: 29.03% Mismatches: 70
Query Match: 11.08% Indels: 67
DB: 10 Gaps: 11

US-09-931-704-1 (1-797) x O22514 (1-326)

```
QY 7 GCTTCGCGAGCGCGGCTCGCTCCCTCCACCTCCGCGAGCTCCG-----GGAGAGGAG 60
DB 152 AlaThrProThrProArgProSerSerProThrSerThrProThrAspGlyGluThr 171
QY 61 CCGCACCAGCGCGCG-----CCAGCCCGAGCCCGGACCTCCGAGCAGGGGACT 111
DB 172 ProArgProGlyProArgSerProSerPro-GlyProProSer----- 188
QY 112 CGTGGGGGATGTTAGCGTGCCTGTGCACGCTCTGGCACCTCCCTGTCAGTGCAGCTC 171
DB 189 -----CysSerProSerProLysSerProSerProProAlaSerSerProPr 204
QY 172 TCATTCGACAGGAGCCAGCGCTGGCCCTCCATCCAGAAAACCTATGACCTCACC 231
DB 204 oArgSerArgProGlyProProAspTyrThrThrSerProSerProProThrProArgSe 224
QY 232 GCTACTGGAGCACCACCTCCGAGCTTGGCTGGGACCTATCTGAACCTACTGGGCCCC 291
DB 224 rValPro-----ProThrProProAlaSer-----ProSerProPr 236
QY 292 CTTTCAACGAGCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCA 351
DB 236 O-----ThrAlaLysProSerProProSerArgGlySerSer----- 248
QY 352 CTGTTGACTTGGAGTGGGAGGCTCAATGACAAACTGCGGCTGACCCAGAACTACG 411
DB 249 -----ProSerProProThrSerProThrProThrProAr 260
QY 412 AGGCTACAGCACCCTTCTGTG----- 433
DB 260 g-----ProProSerTyrSerProSerProThrProProSerSerArgProSerPr 277
QY 434 -----TTACTTTCGCTGGGCTCAACCGTCAAGCTGCGCTGCGCTGAGTGC 477
DB 277 oProLeuArgSerProIleLeuThrProSerProAla-----AlaVa 292
QY 478 GCGCAGCCTGCGCCACTTCTGCACCGAGCTCCAGGCGCTGCTGGGAGCAGCATTCGCGGCG 537
DB 292 lPro-----ProIleGlyArgSerProProSerProIleAspProProCysSerSe 309
QY 538 TCATGGCAGCTTGGGCTACCCACTGCCCGCGCTGCTGGGACT 584
DB 309 rProGluProSerSerProProThrSerPro-ProThrProSerSer 324
```

RESULT 15

```
Q69145 PRELIMINARY; PRT; 327 AA.
ID Q69145;
AC Q69145;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE The first atg start codon is the AA before the stop codon in ORF1
DE (Fragment).
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=P3HR-1;
RX MEDLINE=87198876; PubMed=3033277;
RA Jensen H.B., Farrell P.J., Miller G.;
```

RT "Sequences of the Epstein-Barr Virus (EBV) large internal repeat form
RT the center of a 16-kilobase-pair palindrome of EBV (P3HR-1)
RT heterogeneous DNA [published erratum appears in J Virol 1987
RT Sep;61(9):2950]."; 1506(1987).
RL J. Virol. 61:1495-1506(1987).
DR EMBL; M15972; AAA66541.1; -.
FT NON TER 1
SQ SEQUENCE 327 AA; 33982 MW; 84EAPF95D4ED97A CRC64;

Alignment Scores:

Pred. No.:	0.000216	Length:	327
Score:	161.50	Matches:	81
Percent Similarity:	37.07%	Conservative:	15
Best Local Similarity:	31.27%	Mismatches:	72
Query Match:	11.08%	Indels:	91
DB:	12	Gaps:	18

US-09-931-704-1 (1-797) x Q69145 (1-327)

Qy	13	CCGGAGCGCGGCTCCCTCCACTCCGCCAGCCT-----	48
Db	10	ProAspProGlySerAlaSerProAlaAspArgProHisSerGlyArgLeuLeuGlyAla	29
Qy	49	-----CCGGAGAGGAGCGCCGCCCGCGGC	75
Db	30	SerArgArgGlyTyrPheCysProSerLeuCysProSerGluGluProGlyThrSerGly	49
Qy	76	CCAGCCCGAGCCCGCTCCGCTCCGAGCAGG-----GGACTCGTGGGGGATGTTAG	126
Db	50	-ThrProGluProLeuGlyProAlaSerArgProGlyLeuArgSerProLeuSe	69
Qy	127	CCTGCTCTGCAC-----GGTGTCTCGCACCTCC	156
Db	69	r---ProValysProLysGluCysLeuArgGlyAlaThrLeuGlyAlaGlnAlaPro---	87
Qy	157	CTGCAGTCCAGCTCTCAATGCAC-----AG	183
Db	88	---GluSerArgGlyGlnGlyHisLeuArgValProProArgValProGlyGlnProG1	106
Qy	184	GGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATACCTACCGCTACCTGGAGC	243
Db	106	uglyProArgGln---ProGlyArgProGlnArgProValProArgProPheProGlyue	125
Qy	244	ACCAACTCCG-----CAGCTTGGCTGGGACCTATCTGAACCTACCTGG	285
Db	125	uGlnSerProGlyCysProProGluGlyThrLeuGly-----ValProSe	140
Qy	286	GGCCCTTTTCAACGAGCGAGACTTCAACCTCCCGCGCTGGGGCGAGAGACTCTGCCCA	345
Db	140	rProProLeuGln---AlaArgAlaSerProSerArgGlyAlaSer-----	155
Qy	346	GGGCCACTTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAACTGGCGCTGACCCAGA	405
Db	156	-----LeuGly-----ProHisValGlnProHisArgAspProSe	167
Qy	406	ACTAGAGCGCTACAGCCACTTCTGTGTACTTGGTGGCTCAACCGCTCAGGCTGCCA	465
Db	167	r-----GlyProAspProThr-----GlyProSerLeuCysPr	179
Qy	466	CTGCTGAGCTGGCGG-----CAGCTGGCCCACTTCTGCACAGCCTCC	510
Db	179	o-----ProAlaProLeuGlnProSerLeuHisProArgProGlnLeuLeuAlaSerPr	197
Qy	511	AGGGCTGCTGGGGAGCATTTGGGG-----CGTCATGGCAGCTCTGGGCTACCCACTGC	564
Db	197	oGlyProProGlyGlnProGluGlyProArgGlnProGly-ArgValAlaPheProLeuP	217
Qy	565	CCAGCCG---CTGCTGGGAGTGAACCCACTGGACTCTGCGCCCTGCCAC	614
Db	217	roTriProLeuLeuProAlaSerHisProSerProLeuSerLeuProProHis	234

Search completed: January 27, 2003, 16:16:41

Job time : 59.8794 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:43:12 ; Search time 33.5304 Seconds
(without alignments)
6334.602 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaagcttcgacggacc.....tctctcttcgtccccc 797

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cpn2.1/USPTO.spool/US09931704/runat.27012003.154125.3585/app.query.fasta_1.7189
-DB=A_Geneseq_101002 -QPMF=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704 @CGN 1.1.222 @runat.27012003.154125.3585 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: A_Geneseq_101002.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1364	93.6	253	22	AA25831	Human protein sequ
2	1226	84.1	225	19	AA29715	Human neurotrophic
3	1226	84.1	225	19	AA56141	Amino acid sequenc
4	1226	84.1	225	20	AA94466	Human cardiostroph
5	1226	84.1	225	21	AA87813	Human NNT-1 protei
6	1226	84.1	225	22	AA63543	Amino acid sequenc
7	1226	84.1	225	23	AAU78176	Human novel neurot
8	1226	84.1	225	22	AB11896	Human cardiostroph
9	1225	84.0	321	22	AA79399	Human protein SEQ
10	1218	83.5	260	22	AA78415	Human cardiostroph
11	1214	83.3	223	22	AAE0828	Human cardiostroph
12	1193	81.8	225	19	AA29716	Mouse neurotrophic
13	1193	81.8	225	19	AA56142	Amino acid sequenc
14	1193	81.8	225	21	AA87814	Murine NNT-1 prote
15	1193	81.8	225	23	AAU78177	Mouse novel neurot
16	1169	80.2	215	21	AA819586	Human interleukin-
17	1136	77.9	215	21	AA819587	Mouse interleukin-
18	885	60.7	164	22	AB840317	Peptide #7823 enco
19	885	60.7	164	22	AB24716	Protein #6715 enco
20	885	60.7	164	22	AA61118	Human brain expres
21	885	60.7	164	22	AA73827	Human bone marrow
22	885	60.7	164	22	AA20115	Peptide #6549 enco
23	885	60.7	164	22	AA34012	Peptide #8049 enco
24	885	60.7	164	23	AB843716	Human peptide enco
25	166.5	11.4	1093	14	AA41001	Human myotonic dys
26	163.5	11.2	208	20	AA709197	Human DNAX interle
27	163	11.2	208	20	AA709196	Human DNAX interle
28	160	11.0	1064	17	AA93254	Collagen-like poly
29	160	11.0	1064	17	AA957652	Collagen-like poly
30	160	11.0	1065	14	AA37741	Collagen-like poly
31	159	10.9	446	22	AB70063	Drosophila melanog
32	159	10.9	1690	22	AA23916	Human EST encoded
33	158	10.8	574	22	AA52322	WASP homolog prote
34	158	10.8	574	22	AA67370	Amino acid sequenc
35	157.5	10.8	330	19	AA57645	Collagen-like poly
36	157.5	10.8	408	17	AA07539	Collagen like prot
37	157.5	10.8	408	22	AA72737	Repetitive protein
38	157.5	10.8	408	22	AA64007	CLP protein sequen
39	157.5	10.8	489	22	ABG14285	Novel human diagno
40	154.5	10.6	1838	15	AA53257	Human collagen (Iy
41	154	10.6	1023	23	AA82954	Human homologue of
42	154	10.6	1061	20	AA87504	Human N-methyl-D-a
43	153	10.5	318	22	ABG19764	Novel human diagno
C 44	152	10.5	531	20	AA40098	Spider silk protei
C 45	152	10.5	595	12	AA814309	N.clavipes draglin

ALIGNMENTS

RESULT 1
AA25831
ID AA25831 standard; Protein; 253 AA.
XX
AC AA25831;
XX
DT 16-OCT-2001 (first entry)
DE
DE Human protein sequence SEQ ID NO:1346.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antasthmatic; antidiabetic; cycostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression, Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder.

Homo sapiens.

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

23-DEC-1999; 99US-0471275.

21-JAN-2000; 2000US-0488723;
25-APR-2000; 2000US-0552317.

(HYSE-) HYSEQ INC.

Tang YT, Liu C,

WPI; 2001-457603/49.

DATE: 11/11/74

isolated human po
treatment and dia

Claim 20; Page 278; 1217pp; English.

AAH99166 to AAH99904 encode the huma

cells they are expressed in, such as: anti-inflammatory; anti-neoplastic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; viricide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerrary; antiulcer; osteopathic; dermatological; antiallergic; anti-laschmatic; anti-diabetic; cytostatic; neuroprotective; antidepressant; nootropic; anti-parkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

Sequence 253 AA;

Assignment Scores:

Length:	2.41e-104	253
Matches:	1364.00	251
Conservative:	99.21%	0
Mismatches:	99.21%	2
Indels:	93.55%	0
Gaps:	22	0

9-931-704-1 (1-797) x AAM25831 (1-253)

6 AGCTTCGCCGGAGCCGCGGCTCGCCCTCCCACTCCGCCAGCCTCCGGAGAGGAGCCGCA 65

1 SerPheAlaGlyAlaAlaAlaAArgProSerThrProProAlaSerGlyArgGlyAlaAla 20

66 CCGGCCGCCAGCCCCAGCCCCATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTA 125

21 ProGlyArgProGlyProSerProMetAspLeuArgAlaGlyAspSerTrpGlyMetLeu 40

QY	126	GCCTGCTGTGACAGTGTCTCTGGCACTCTCCCTGCAGTGCACAGCTCTCAATCGCACAGG	180
Db	41	AlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsnArgThrGly	60
QY	186	GACCCAGGCGCTGGCCCTCCATCCAGAAAACTATGACCTCACCCGGCTACCTGGAGCAC	245
Db	61	AspProGlyProGlyProSerIleGlnysThrTyrAspLeuThrArgTyrLeuGluHis	80
QY	246	CAACTCCGACAGCTGTGCTGGGACCTATCTGAACCTACTGGGGCCCCCTTTCAACGAGCCA	305
Db	81	GlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGluPro	100
QY	306	GACTTCAACCCCTCCCGCTGGGGCAGACAGACTCGCCAGGGCCACTGTTCTACCTTGGAG	365
Db	101	AspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGlu	120
QY	366	GTCTGGCGAAGCCTCAATCACAACTCGCGGCTGACCCAGAACCTACGAGGCGCTACAGCCAC	425
Db	121	ValTrpArgSerLeuAsnAspIysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHis	140
QY	426	CTTCTGTGTATCTTGCTGGGCTCAACCGCTCAGGCTGCCACTGTGTAGCTGGCGCGCAGC	485
Db	141	LeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgSer	160
QY	486	CTGGCCCACTTGTGACACAGCTCTCAGGCGCTCTGGGCAGCATTGCGGCGCTCATGGCA	545
Db	161	LeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAla	180
QY	546	GCTCTGGGGTACCACCTGCGCCCGCCGCTGCGTGGGACTGAACCCACTTTGGACTCTCTGC	605
Db	181	AlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGly	200
QY	606	CTTGCCCAAGTGACTTCTCCAGAAAGATGGACGACTTCTGGCTGCTGAAGAGAGCTGCAG	665
Db	201	ProAlaHisSerAspPheLeuGlnIlysMetAspAspPheTrpLeuLeuIysGluLeuGln	220
QY	666	ACCTGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAGAAGATGCAGCTCCA	725
Db	221	ThrTrpLeuTrpArgSerAlaIysAspPheAsnArgLeuIysIleIysMetGlnProPro	240
QY	726	GCAGCTGCAGTCAACCTGCACCTGGGGGCTCATGGCTC	764
Db	241	AlaAlaAlaValThrLeuHisLeuGluValaHisGlyPhe	253

RESULT 2

RESULT 2
AAW29715
ID AAW29715 standard: Protein: 225 AA.

AA
AC AAW29715;

DT 09-NOV-1998 (first entry)

Human neurotrophic factor NNT-1.

NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant; Alzheimer's disease; Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; peripheral neuropathy; dystrophy; neural retina degeneration; common variable immunodeficiency; CVID; selective IgA deficiency; hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic therapy.

OS Homo sapiens.

XX	Key	Location/Qualifiers
FH		

FT Peptide

FT /label= sig_peptide

FT	Protein	28..225
FT	Protein	28..225

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

[illegible]

XX 02-FEB-1998; 98WO-US02363.
PF XX
PR 30-JAN-1998; 98US-0016534.
PR 03-FEB-1997; 97US-0792019.
XX
XX
PA (AMGE-) AMGEN INC.
PA
XX
XX Chang M, Elliot GS, Sarmiento U, Senaldi G;
PI
XX WPI; 1998-437475/37.
DR N-PSDB; AAV47510-11.
DR
XX
XX Newly isolated nucleic acid encoding human or murine neurotrophic
PT factor NNT-1 - useful for treatment of neurological and
PT immunological diseases or inflammation, also as vaccine adjuvant
XX
XX Claim 12; Fig 3; 120pp; English.
PS
XX
XX This is the amino acid sequence of a novel neurotrophic factor,
CC designated NNT-1, that is a growth factor for neurons and for B or
CC T cells. It was deduced from isolated cDNA (see AAV47510) and
CC genomic DNA (see AAV47511) clones. Vectors containing the cDNA or
CC genomic DNA and host cells are provided for use in the production
CC of NNT-1 polypeptides. These are used to treat: (i) neurological
CC or immunological diseases, specifically Alzheimer's, Parkinson's
CC or Huntington's diseases, amyotrophic lateral sclerosis,
CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
CC degeneration of the neural retina, or conditions characterised by T
CC or B cell defects, e.g. common variable immunodeficiency (CVID),
CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
CC inflammation. NNT-1 is also able to boost immunoreactivity and
CC antibody production following vaccination, and, since it inhibits
CC tumour necrosis factor production, it may also be useful for
CC treating sepsis. In addition, cells that have been engineered to
CC express NNT-1 can be implanted, or nucleic acids are delivered in
CC gene therapy vectors.
XX
SQ Sequence 225 AA;

Alignment Scores:
Pred. No.: 5,786-93 Length: 225
Score: 1226.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW29715 (1-225)

QY 90 ATGGACCTCGAGCAGGGGACTGCTGGGGATGCTTAGCGTCTGTCAGCGGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
QY 210 CAGAAACCTATGACTTCACTCCCTACCTGAGCAGACCACTCCGCGAGCTTGGTGGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGACTTACTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCCCTGGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 330 GCAGAGACTCTGCCAGGGCCACTGTGTACTTGGAGTGTGGCGAAGCCCTCAATGACAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
QY 390 CTCGGCTGACCCAGAACCTACGAGGCTACAGGCTACAGCACTTCTGTGTACTTGGCTGCTC 449

Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCAGGCTGCCACTGCTGAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGGCTCTGCTGGCAGCATTCGGCGCTCATGGCAGCTCTGGGCTACCCACTGCCCCAG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
QY 570 CCGTCTCGCTGGGACTGAACCCACTTGGACTCTGCTGGCCCTGCCACAGTCTCTCCAG 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGGAGCACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGGCGCTGGCGCAAG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGCTCAAGAGATGCAGCTCCAGCAGCTGCAGTGCAGTGCAGTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GlyAlaHisGlyPhe 225
RESULT 3
AAW56141
ID AAW56141 standard; Protein; 225 AA.
XX
AC AAW56141;
DT 13-JUL-1998 (first entry)
XX
DE Amino acid sequence of human neurotrophic factor NNT-1.
XX
KW Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /note= "signal peptide"
FT Protein 28..225
FT Protein /note= "mature protein"
XX
PN US5741772-A.
XX
PD 21-APR-1998.
XX
PF 03-FEB-1997; 97US-0792019.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 1998-260526/23.
DR N-PSDB; AAV22652.
XX
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
PT useful for stimulating growth of motor and sympathetic neurons
XX
PS Claim 1; Fig 3; 41pp; English.
XX
CC The present sequence represents a human neurotrophic factor, designated
CC NNT-1, which is capable of stimulating growth of motor or sympathetic
CC neurons. The NNT-1 protein is useful in the treatment of neurological
CC diseases characterised by the degeneration and death of particular
CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimers's disease,
 CC stroke and various degenerative disorders affecting vision.
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 5,78e-93 Length: 225
 Score: 1226.00 Matches: 225
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 84.09% Indels: 0
 DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW56141 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCTGTGCAGCGTGCTCTGG 149
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 150 CACTCCCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCTGGCCCTCCATC 209
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
 QY 210 CAGAAACCTATGACCTCACCCTACCTGGAGCACCACCTCCGAGCTGGCTGGGACC 269
 Db 41 GlnLysThrTrpAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 270 TATCTGAACCTACCTGGGCCCCCTTCAACGAGCCAGACTTCAACCTCCCGCCTGGGG 329
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 330 GCAGAGACTGCGCAGGCCCACTTTGACTTGGAGGTGTGGCAAGCCTCAATGACAAA 389
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
 QY 390 CTGGCGCTGACCGAGACTAGAGCCCTACAGCCACCTCTGTGTACTTGGTGGCCCTC 449
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 QY 450 AACCGTCAAGGCTGCGCAGCTGAGCTGGCGCGCAGCCTGGCCCACTTTCGACACCGCCTC 509
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 510 CAGGCGCTGCGGCGAGCATTTGCGGGCGTCAATGGCAGCTCTGGGTACCCACTGCCCCAG 569
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
 QY 570 CCGTGTCTGGGACTGAACCCACTTGGACTCTGGCCCTGGCCACAGTCACTTCTCCAG 629
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
 QY 630 AAGATGGACGACTTCTGGCTGCTGAAGGAGCTGCAGACTGGCTGTGGCGCTCGGCCAAG 689
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 690 GACTTCAACCGCTCAAGAGATGAGCTCCAGAGCTCCAGAGCTGAGTCACTGACCTGACCTG 749
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220
 QY 750 GGGGCTCATGCTTC 764
 Db 221 GlyAlaHisGlyPhe 225

RESULT 4
 AAW94466
 ID AAW94466 standard; Protein; 225 AA.
 XX
 AC AAW94466;
 XX
 DT 22-APR-1999 (first entry)
 XX
 DE Human cardiotrophin-like cytokine protein.
 XX

KW Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;
 KW CLC; IL-6; diagnosis; detection; immune system-related disorder;
 KW cancer; cardiac disorder; heart failure; hypertension; cancer;
 KW autoimmune disorder; infection.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..27 /label= signal
 FT Protein 28..225 /label= Cardiotrophin-like_cytokine
 FT Domain 74..79 /label= CD-I
 FT /note= "conserved domain"
 FT Domain 150..156 /label= CD-II
 FT /note= "conserved domain"
 FT Domain 194..198 /label= CD-III
 FT /note= "conserved domain"
 XX
 PN WO9900415-A1.
 XX
 PD 07-JAN-1999.
 XX
 PF 29-JUN-1998; 98WO-US13129.
 XX
 PR 30-JUN-1997; 97US-0051311.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Shi Y;
 XX
 DR WPI: 1999-095678/08.
 DR N-PSDB; AAX16161.
 XX
 PT New isolated cardiotrophin-like cytokine nucleic acid - used to
 PT develop products for treating cardiac and immune system disorders,
 PT e.g. heart failure, hypertension, cancers, autoimmune disorders and
 PT infections
 XX
 PS Claim 1; Fig 1; 103pp; English.
 XX
 CC The present invention relates to a novel cardiotrophin-like cytokine
 CC (CLC) protein which is a member of the interleukin 6 (IL-6) cytokine
 CC family. The present sequence represents the human CLC protein. The
 CC present invention also describes screening methods for identifying
 CC agonists and antagonists of CLC activity, as well as methods for
 CC detecting cardiac and immune system-related disorders and
 CC therapeutic methods for treating cardiac and immune system-related
 CC disorders, e.g heart failure, hypertension, cancers, autoimmune
 CC disorders and infections.
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 5,78e-93 Length: 225
 Score: 1226.00 Matches: 225
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 84.09% Indels: 0
 DB: 20 Gaps: 0

US-09-931-704-1 (1-797) x AAW94466 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTGCTGTGCAGCGTGCTCTGG 149
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 150 CACTCCCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCTGGCCCTCCATC 209
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 210 CAGAAACCTATGACCTCACCGCTACCTGAGCAGCACTCCGAGCTTGGCTGGACC 269
DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGG 329
DB 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 330 CGAGAGACTGCCCCAGGCGCACTTGTGCTGGAGGTGGCGAGCTCAATGACAAA 389
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
QY 390 CTGCGGCTGACCCAGCACTACAGGCGCTACAGCCACTTCTGTACTTGGCTGGCCTC 449
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCTCAGGCTGCCACTCTGCTGAGTGGCGCCGAGCTGGCCACTTTCGACCCGCTC 509
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGCGCTGCGGAGCAATTCGGGCGCTCATGGAGCTCTGGGCTACCCACTGCCCGAG 569
DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
QY 570 CGCTGCTGCTGGAGCTGAACCCACTTGGACTCTCTGGCCCTGCGCCAGCTGCTCCAG 629
DB 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGGAGCACTTCTGGCTGTGAAGAGCTGAGAGCTGCGAGCTCTGGGCTACCCACTGCCCGAAG 689
DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGGCTCAAGAGAGATGAGCTGAGCTGCGAGCTGCGAGCTGCGAGCTG 749
DB 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
DB 221 GlyAlaHisGlyPhe 225

RESULT 5
AAY87813
ID AAY87813 standard; Protein; 225 AA.
AC AAY87813;
XX
DT 24-AUG-2000 (first entry)
XX
DE Human NNT-1 protein.
XX
KW NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;
KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
KW myotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
KW retinopathy; immune disorder; hematopoietic disorder.
XX
OS Homo sapiens.
XX
FN US6054294-A.
XX
PD 25-APR-2000.
XX
PF 12-DEC-1997; 97US-0988819.
XX
PR 03-FEB-1997; 97US-0792019.
XX
PA (AMGE-) AMGEN INC.
XX
PI Chang M;
XX
DR WPI; 2000-338492/29.

DR N-PSDB; AAA39481.

XX New nucleic acids encoding neurotrophic factors useful for stimulating growth of motor or sympathetic neurons for treating neuron cell damage

XX Claim 1c; Fig 3; 42pp; English.

XX This invention describes a novel nucleic acid molecule (I) encoding a novel neurotrophic factor (NNT-1) (II) which has neurotrophic, neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and ophthalmological activity. (I) is useful for producing NNT-1 polypeptides which are useful for treating patients in whom various cells of the central, autonomic, or peripheral nervous system have degenerated and/or have been damaged by congenital disease, trauma, mechanical damage, surgery, stroke, ischemia, infection, metabolic disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1 proteins are used to treat diseases like Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's disease, peripheral neuropathy induced by diabetes or other metabolic disorders, and/or dystrophies or degeneration of the neural retina such as retinitis pigmentosa, drug-induced retinopathies, stationary forms of night blindness, progressive cone-rod degeneration, immune disorders and hematopoietic disorders. (II) is effective in treating neurological conditions and promotes neuron regeneration. Neural functions are effectively restored in patients suffering from various neurological disorders. This sequence represents the human NNT-1 protein described in the method of the invention.

SQ Sequence 225 AA;

Alignment Scores:
Pred. No.: 5,78e-93 Length: 225
Score: 1226.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 21 Gaps: 0

US-09-931-704-1 (1-797) x AAY87813 (1-225)

QY 90 ATGAGACTCCGAGCAGGGGACTCTGCGGGGATGTAGCGTGTGTCAGCGTCTCTGG 149
DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCAGAGGGAGCCAGGGCTCGCCCTCCATC 209
DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
QY 210 CAGAAACCTATGACCTCACCGCTACCTGGAGCAGCACTCCGAGCTTGGCTGGGACC 269
DB 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 329
DB 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 330 CGAGAGACTCTGCCAGGCGCACTTGTGACTTGGAGGTGGCGAGCTCAATGACAAA 389
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
QY 390 CTGCGGCTGACCCAGCACTACAGGCGCTACAGCCACTTCTGTGTTACTTGGCTGGCCTC 449
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCTCAGGCTGCCACTCTGCTGAGTGGCGCCGAGCTGGCCACTTTCGACCCGCTC 509
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGCGCTGCGGAGCAATTCGGGCGCTCATGGAGCTCTGGGCTACCCACTGCCCGAG 569
DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160

QY 570 CCGCTGGCTGGAGTCAACCCACTTGGACTCCTGGCCCTGCCAGTCACTCTCTCCAG 629
 |||||
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
 |||||
 QY 630 AAGATGACGACTTCTGGCTGCTGAAGAGGTGCGAGCTGGCTGTGGCTGGCCCAAG 689
 |||||
 Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 |||||
 QY 690 GACTTCAACCGCTCAAGAGAAGATGCGAGCTCCAGAGTGGAGTCACTCTGCACCTG 749
 |||||
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
 |||||
 QY 750 GGGGCTCATGGCTTC 764
 |||||
 Db 221 GlyAlaHisGlyPhe 225
 |||||
 RESULT 6
 AAG63543
 ID AAG63543 standard; Protein; 225 AA.
 XX AAG63543;
 AC AAG63543;
 XX
 DT 15-OCT-2001 (first entry)
 DE Amino acid sequence of a human NNT-1 protein.
 XX
 KW NNT-1; CLF-1; sCNTFRalpha; nervous system; neuron; nervous system;
 KW neuro-muscular function; tumour; immune system; haematopoietic system;
 KW reproductive system; liver; skeletal muscle; neurodegenerative disease;
 KW amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
 KW muscular mass; paralysis; cancer; obesity; fertility; endometriosis;
 KW blastocyst implantation; thrombosis; retinal disease;
 KW retinal pigmentosis.
 XX
 KW Homo sapiens.
 OS
 XX
 PN WO200155172-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 26-JAN-2001; 2001WO-FR00253.
 PF
 XX 27-JAN-2000; 2000FR-0001035.
 PR
 PR 12-OCT-2000; 2000FR-0013089.
 XX
 XX (FABR) FABRE MEDICAMENT SA PIERRE.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA
 XX Elson G, Gauchat J, Plun-Favreau H, Chevalier S, Gascan H;
 PI
 DR WPI; 2001-488773/53.
 DR N-PSDB; AAH74484.
 DR
 XX A complex comprising a NNT-1 protein and a CLF-1 and/or sCNTFRalpha
 PT protein useful to treat neurodegenerative disease including Parkinson's
 PT and Huntington's, obesity and cancer -
 XX
 XX Claim 2; Page 58; 67pp; French.
 PS
 XX The present sequence represents a human NNT-1 protein. The specification
 CC describes a complex comprising a NNT-1 protein and a CLF-1 and/or
 CC sCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate
 CC activity of the sCNTFRalpha/gp130/LiFrbeta receptor complex, or to
 CC induce phosphorylation of the tyrosine of gp130 and LiFrbeta,
 CC particularly where cells expressing the receptor complex are in the
 CC central or peripheral nervous system, in neurons implicated in
 CC neuro-muscular function or in skeletal muscle. The complex or
 CC antibodies are also used to decrease the survival, growth or
 CC proliferation of tumour cells or to facilitate the proliferation and/or
 CC inhibit differentiation of cells stocks. The complex is also used to
 CC modulate activity of the gp130/LiFrbeta receptor or cells expressing
 CC that receptor, particularly those cells implicated in the immune,
 CC haematopoietic, nervous or reproductive system, the liver or skeletal

CC muscle. Molecules of the invention may be used to prevent or treat
 CC neurodegenerative diseases including amyotrophic lateral sclerosis,
 CC Parkinson's and Huntington's disease, to repair or regenerate nervous
 CC or muscular tissue or to maintain muscular mass in paralysis patients.
 CC They may also be used to treat cancer, obesity and associated diseases,
 CC and to improve fertility, particularly to avoid endometriosis and/or
 CC assist blastocyst implantation, thrombosis, or retinal disease,
 CC particulate retinal pigmentosis.

XX Sequence 225 AA;

Alignment Scores:
 Pred. No.: 5,78e-93 Length: 225
 Score: 1226.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 84.09% Indels: 0
 DB: 22 Gaps: 0

US-09-931-704-1 (1-797) x AAG63543 (1-225)

QY 90 ATGGACCTCCGAGAGGAGTCTGCTGGGGATGTTAGCTGTGCTGTGCAGGTCTCTGG 149
 |||||
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 |||||
 QY 150 CACCTCCCTGCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209
 |||||
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
 |||||
 QY 210 CAGAAAACCTATGACTACCCGCTACCTGGAGCACAACTCCGAGCTGGTGGGACC 269
 |||||
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 |||||
 QY 270 TATCTGAATCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGG 329
 |||||
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
 |||||
 QY 330 GCAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGGAGAGCCCTCAATGACAAA 389
 |||||
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
 |||||
 QY 390 CTGGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGTGGTGGCTC 449
 |||||
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 |||||
 QY 450 AACCGTCAGGCTGCCACTGTGAGCTGCGCGCAGCTGCGCCACTTCTGCACAGCCTC 509
 |||||
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
 |||||
 QY 510 CAGGGCTGCTGGGAGCATTTGCGGCGTCATGCGAGCTCTGGGCTACCCACTGCCCCAG 569
 |||||
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
 |||||
 QY 570 CGCTGCTCGGGAGTCAACCCACTTGGAGTCTCTGGGCGCTGCCACAGTACTTCTCCAG 629
 |||||
 Db 161 ProLeuProGlyThrGluProThrTrpTrpProGlyProAlaHisSerAspPheLeuGln 180
 |||||
 QY 630 AAGATGAGAGTCTTCTGGCTGCTGAAGAGTGCAGACCTGGCTGTGGCTCGGCTCGGCAAG 689
 |||||
 Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 |||||
 QY 690 GACTTCAACCGCTCAAGAGAAGATGCGAGCTCCAGAGTGGAGTCACTCTGCACCTG 749
 |||||
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
 |||||
 QY 750 GGGGCTCATGGCTTC 764
 |||||
 Db 221 GlyAlaHisGlyPhe 225
 |||||
 RESULT 7
 AAU78176
 ID AAU78176 standard; Protein; 225 AA.
 XX

AC AAU78176;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Human novel neurotrophic factor NNT1.
 XX
 KW Human, NNT1; neurotrophic factor; IGF-related disease;
 KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
 KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;
 KW graft versus host disease; infertility; miscarriage; preterm labour.
 XX
 OS Homo sapiens.
 XX
 XX WO200215977-A2.
 XX
 XX 28-FEB-2002.
 XX
 XX 17-AUG-2001; 2001WO-US25906.
 XX
 XX 18-AUG-2000; 2000US-226436P.
 PR
 XX 16-AUG-2001; 2001US-0931704.
 XX
 XX (AMGE-) AMGEN INC.
 PA
 XX
 XX Senaldi G;
 PI
 XX WPI; 2002-280867/32.
 DR
 XX N-PSDB; ABK11647.
 DR
 XX
 XX Treating Immunoglobulin E-related disease, modulating IgE levels in a
 PT patient, preventing IGE-related disease and treating allergic diseases,
 PT involves administering NNT-1 inhibitor to a patient -
 XX
 XX Claim 2; Fig 3; 63pp; English.
 PS
 XX The invention relates to treating Immunoglobulin E (IGE)-related disease,
 CC modulating IGE levels in a patient, preventing an IGE-related disease,
 CC and treating allergic diseases, comprising administering a
 CC therapeutically effective amount of novel neurotrophic factor (NNT)-1
 CC inhibitor to a patient. Also included are a method of diagnosing an
 CC IGE-related disease or susceptibility to an IGE-related disease, by
 CC determining the presence or amount of expression of an NNT1 polypeptide
 CC encoded by a NNT1 nucleotide sequence, its fragment or naturally
 CC occurring variant, and diagnosing an IGE-related disease or
 CC susceptibility of an IGE-related disease based on the presence or amount
 CC of expression of the polypeptide and a pharmaceutical composition for use
 CC in treating IGE-related disease, comprising the NNT1 inhibitor.
 CC The NNT1 inhibitor is useful for preventing and treating IGE-related
 CC disease, modulating IGE levels, and treating allergic diseases e.g.
 CC Type I allergic disease, allergic rhinitis, eczema, dermatitis,
 CC pollinosis, asthma, immune diseases and disorders, diseases involving
 CC abnormal cell proliferation including cancer, arteriosclerosis and
 CC vascular restenosis, diseases and conditions relating to dysfunction of
 CC immune system including rheumatoid arthritis, psoriatic arthritis,
 CC inflammatory arthritis, osteoarthritis, inflammatory joint disease,
 CC autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory
 CC bowel disease, transplant rejection, and graft versus host disease, and
 CC reproductive diseases and disorders including infertility, miscarriage,
 CC preterm labour and delivery, and endometriosis. The present sequence
 CC represents human NNT1.
 XX
 SQ Sequence 225 AA;
 Alignment Scores:
 Pred. No.: 5,78e-93 Length: 225
 Score: 1226.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 84.09% Indels: 0

DB: 23 Gaps: 0
 US-09-931-704-1 (1-797) x AAU78176 (1-225)
 QY 90 ATGACCTCCAGAGGAGGAGGACTCGTGGGGGATGTAGCGTGTGACGGTGTCTCTGG 149
 DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 150 CACCTCCCTGAGTGCCAGCTCTCAATCGCACAGGGGACCCAGGGGCTTGGCCCTCCATC 209
 DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
 QY 210 CAGAAAACCTATGACCTCACCCGCTACCTGGAGACCAACTCCGAGTGTGGTGGACC 269
 DB 41 GlnLysThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 270 TATCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 329
 DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 330 GCAGAGACTCTGCCAGGGCCACTGTTCACTTGGAGGTGTGGGAGGCTCAATGACAAA 389
 DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
 QY 390 CTGGGCTGACCCAGAACTACAGGCTTACAGCCACTCTGTGTACTTGTGGTGGCTC 449
 DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 QY 450 AACCGTCAAGGTGCCACTGCTGAGCTGGCGGCGAGCTGGCCACTTTCGACACGACCTC 509
 DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 510 CAGGCGCTGCTGGGCGAGCATTTGCGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAG 569
 DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
 QY 570 CCGTGTCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCCAAGTACTTCTCCAG 629
 DB 161 ProLeuProGlyThrGluProThrTrpProGlyProAlaHisSerAspPheLeuGln 180
 QY 630 AAGTGGAGGACTTCTGCTGCTGAGGAGCTGAGAGCTGGCTGGCTGGCTGGCTGGCTG 689
 DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 690 GACTTCAACCGCTCAAGAAGATGAGCCCTCCAGAGCTGCAGTGCAGTGCAGTGCAGCTG 749
 DB 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220
 QY 750 GGGGCTCATGGCTTC 764
 DB 221 GlyAlaHisGlyPhe 225
 RESULT 8
 ABBI1896
 ID ABBI1896 standard; peptide; 321 AA.
 XX
 AC ABBI1896;
 XX
 DT 11-JAN-2002 (first entry)
 DE Human cardiotrophin-like cytokine homologue, SEQ ID NO:2266.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;

KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.

OS Homo sapiens.

XX WO200157188-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03800.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

XX N-PSDB; ABA09140.

XX Human proteins and DNA encoding sequences useful for preventing,

XX e.g. arthritis and cancer -

XX Claim 20; Page 273; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
invention also relates to vectors and recombinant host cells comprising a
nucleotide of the invention, methods of producing the novel polypeptides,
antibodies against the polypeptides, methods of detecting the nucleotides
or polypeptides in a sample, and methods of identifying compounds which
bind to polypeptides of the invention. Although novel, many of the
polypeptides of the invention have homology to known proteins, thereby
giving an insight into their probable biological activities, and hence
potential therapeutic applications. The polypeptides of the invention may
have various activities, including cytokine, cell proliferation or cell
differentiation activities; stem cell growth factor activity;

XX haematopoiesis regulatory activity; tissue growth activity;
immunomodulatory activity; activin- or inhibin-related activities;
chemotactic or chemokinetic activities; haemostatic, thrombotic or
thrombolytic activities; receptor or ligand activities; or may be
involved in oncogenesis, cancer cell proliferation or metastasis.
XX depending on their biological activities, polypeptides and nucleotides of
the invention are useful for preventing, treating or ameliorating medical
conditions, e.g., by protein or gene therapy. Such conditions include
cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
proliferative retinopathy, atherosclerosis, coronary heart disease,
arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
vascular growth. Polypeptides involved with tissue regeneration and
repair (or nucleic acids encoding them) may be used to promote wound
healing (e.g., of burns, incisions and ulcers), while those with
immunomodulatory activities may be used in the treatment of viral,
bacterial and fungal infections in addition to immune disorders.
XX Polypeptides with growth factor activity may be used in cell cultures to
promote cell growth. For example, such polypeptides may be used to
manipulate stem cells in culture to give rise to neuroepithelial cells
that can be used to augment or replace cells damaged by illness,
XX autoimmune disease or accidental damage. The polypeptides and nucleotides
may also be used in the diagnosis of the above conditions, and in drug
screening techniques. The present sequence represents a novel human
polypeptide of the invention.

XX Sequence 321 AA;

XX Alignment Scores:

XX Pred. No.: 7,65e-93 Length: 321

XX Score: 1225.00 Matches: 232

XX Percent Similarity: 91.02% Conservative: 1

XX Best Local Similarity: 90.62% Mismatches: 11

XX Query Match: 84.02% Indels: 12

DB: 22 Gaps: 3
US-09-931-704-1 (1-797) x ABB11896 (1-321)

Qy 33 CCCACTCCGCGAGCCTCCGGGAGA---GGAGCGGCACCC---GGCGGGCCACCCCGCCAGC 86

Db 66 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 85

Qy 87 CCCATGGAC-----CTCCGAGCAGGGGACTCTGTGG 116

Db 86 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThrGlyAspSerTrp 105

Qy 117 GGGATGTTAGCGTGTGTCACGGTGTCTGGCACCTCCCTGACGTCAGTCAGCTTCAAT 176

Db 106 GlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsn 125

Qy 177 CGCACAGGGGACCCAGGGCTGGCCCTCATCCAGAAAACCTATGACCTCACCCGGTAC 236

Db 126 ArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyr 145

Qy 237 CTGGAGCACCAACTCGCAGCTTGGCTGGGACCTATCTGAACCTACCTGGGCCCTTTC 296

Db 146 LeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPhe 165

Qy 297 AACGAGCCAGACTTCAACCTCCCGCTCCGGGGCAGAGACTCTGCCAGGGCCACTGTT 356

Db 166 AsnGluProAspPheAsnProArgLeuGlyAlaGluThrLeuProAlaThrVal 185

Qy 357 GACTTGGAGGTGGCGAAGCCTCAATGACAACTGGGGCTGACCCAGAGACTAGAGGCC 416

Db 186 AspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAla 205

Qy 417 TACAGCCACTTCTGTGTACTTGGCTGGCTCAACCGTCAGCGTGCCTGCTGAGCTG 476

Db 206 TyrSerHisLeuLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeu 225

Qy 477 CGCCGAGCGCTGGCCCACTTCTGCACAGCCTCCAGGGCTGCTGGCAGCATTTGGGGC 536

Db 226 ArgArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaGly 245

Qy 537 GTCATGGAGCTTGGGCTACCCACTGCCCGCCAGCGCTGCTGGAGCTGACCCACTGG 596

Db 246 ValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrp 265

Qy 597 ACTCTCGGCGCTCCACAGTGTCTTCTCCAGAGATGGACACTTCTGGCTGCTGAAG 656

Db 266 ThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLys 285

Qy 657 GAGCTGCAGACCTGGGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGATG 716

Db 286 GluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMet 305

Qy 717 GAGCTTCAGCAGCTGAGTCACTGACCTGGGCTGAGCTGGGGCTCATGGCTTC 764

Db 306 GlnProProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 321

RESULT 9

AAW79399

ID AAW79399 standard; Protein; 321 AA.

XX AC AAW79399;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3045.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

XX XX

PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PP 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK52532.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 PT
 XX
 PS Claim 20; Page 237; 6221pp; English.
 XX
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAW80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 321 AA;

Alignment Scores:
 Pred. No.: 7,65e-93 Length: 321
 Score: 1225.00 Matches: 232
 Percent Similarity: 91.02% Conservative: 1
 Best Local Similarity: 90.62% Mismatches: 11
 Query Match: 84.02% Indels: 12
 DB: 22 Gaps: 3

US-09-931-704-1 (1-797) x AAW79399 (1-321)

QY 33 CCCACTCCCGAGCTCCGGAGA---GGAGCGCGACCC---GGCGCGCCAGCCCGCAGC 86
 Db 66 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 85
 QY 87 CCCATGGAC-----CTCGAGCAGGGGACTCGTGG 116
 Db 86 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThrGlyAspSerTrp 105
 QY 117 GGGATTAGCTGCTGTGACGGTGCTCTGCACCTCCCTGCAGTCCGAGCTCTCAAT 176
 Db 106 GlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsn 125
 QY 177 CGCACAGGGGACCCAGGCGCTGCCCTCCATCCAGAAACCTATGACCTCACCCCGCTAC 236
 Db 126 ArgThrGlyAspProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyr 145
 QY 237 CTGGAGCACCAACTCCGCGACTTGGCTGGGAGCCTATCTGAACCTACCTGGGCGCCCTTTC 296

Db 146 LeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPhe 165
 QY 237 AACGAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTCCCGAGGCGCACTGT 356
 Db 166 AsnGluProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrVal 185
 QY 357 GACTTGGAGGTGTGGCGAAGCCTCAATGACAACTCCGCGCTGACCCAGAACCTACGAGGCC 416
 Db 186 AspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAla 205
 QY 417 TACAGCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTCAGCTGCTGAGCTG 476
 Db 206 TyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeu 225
 QY 477 CGCGCAGCGCTGGCCCACTTCTGACCCAGCCTCCAGGGCGCTGCTGGCAGCATTGCGGCG 536
 Db 226 ArgArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGly 245
 QY 537 GTCTGGCAGCTCTGGGCTTACCCACTGCCCGCCAGCCGCTGCTGGGACTGAACCCACTGG 596
 Db 246 ValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrp 265
 QY 597 ACTCTGGCGCTGCCACAGTACTTCTCCAGAGATGAGACGACTTCTGGCTGCTGAAG 656
 Db 266 ThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLys 285
 QY 657 GAGCTGCAGACCTGGCTGTGGCGCTCGCGCAAGGACTTCAACCGGCTCAAGAGAGATG 716
 Db 286 GluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMet 305
 QY 717 CAGCCTCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764
 Db 306 GlnProProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 321
 RESULT 10
 AAW78415
 ID AAW78415 standard; Protein; 260 AA.
 XX AAW78415;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1077.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 XX
 FN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX

DR WPI; 2001-476283/51.
 DR N-PSDB; AAK51548.
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 3306; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAW80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX

SQ Sequence 260 AA;

Alignment Scores:
 Pred. No.: 2,74e-92 Length: 260
 Score: 1218.00 Matches: 231
 Percent Similarity: 90.62% Conservativeness: 1
 Best Local Similarity: 90.23% Mismatches: 12
 Query Match: 83.54% Indels: 12
 DB: 22 Gaps: 3

US-09-931-704-1 (1-797) x AAW78415 (1-260)

QY 33 CCCACTCCGCGAGCTCTGGGAGA---GGAGCCGACCC---GGCGGCGCCAGCCCGCCAGC 86
 DB 5 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 24
 QY 87 CCCATGGAC-----CTCCGAGCGAGGACTCGTGG 116
 DB 25 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThrGlyAspSerTrp 44
 QY 117 GGGATGTTAGCGTCTGCTGACGGTGTCTGGACCTCTCCCTGCGAGTCCAGCTCTCAAT 176
 DB 45 GlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsn 64
 QY 177 CGCAGGGGACCCAGGCGCTGGCCCTCCATCCAGAAACCTATGACCTCAGCCGCTAC 236
 DB 65 ArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspProThrArgTyr 84
 QY 237 CTGGAGCACCAACTCCGAGCTTGGCTGGGACCTATCTGAACCTACTGGGCGCCCTTTTC 296
 DB 85 LeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPhe 104
 QY 297 AACGAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGCGCACTGTT 356
 DB 105 AsnGluProAspPheAsnProProArgLeuGlyAlaGluThrLeuProAlaThrVal 124
 QY 357 GACTTGGAGGTGGCGAGCTCAATGACAACTGGCGCTACCCAGCACTACGAGGCC 416
 DB 125 AspLeuGluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAla 144
 QY 417 TACAGCCACCTTCTGTGTACTTGGTGGCCCTCAACCGCTCAGGCTGCCACTGCTGAGCTG 476
 DB 145 TyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeu 164
 QY 477 CGCGCAGCTGGCCCACTTCGACCAAGCTCCAGGCGCTCTGGGCGAGCAFTGGCGGC 536
 DB 165 ArgArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGly 184
 QY 537 GTCATGGAGCTCTGGGCTACCACTGCCCGAGCGCTCCCTGGGACTCAACCCACTTGG 596

DB 185 ValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrp 204
 QY 597 ACTCTGGCGCTGCCACAGTGTCTCCAGAGATGGACGACTTCTGGCTGCTGAAG 656
 DB 205 ThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLys 224
 QY 657 GAGCTGCAGACCTGGCTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAGAGAGATG 716
 DB 225 GluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMet 244
 QY 717 CAGCTCCAGAGCTCCAGTCCAGCTCCCTGCGAGCTGGGGCTCATGGCTTC 764
 DB 245 GlnProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 260

RESULT 11

AAE00828 AAE00828 standard; Protein; 223 AA.

XX AC AAE00828;

DT 02-JUL-2001 (first entry)

XX Human cardiotrophin-like cytokine (CLC) protein.

XX Human; biologically active complex; haemopoietin receptor; NR6;
 KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;
 KW differentiation; cell survival; neurotrophic activity.

OS Homo sapiens.

FX Key Location/Qualifiers

FT Peptide 1..27

FT Protein /label= Signal_peptide

FT Protein 28..223

FT Protein /label= Human mature CLC protein

FT Protein /note= "Cardiotrophin-like cytokine"

XX WO200127157-A1.

PN 19-APR-2001.

PD 06-OCT-2000; 2000WO-AU01216.

PP 08-OCT-1999; 99AU-0003327.

PR 12-MAY-2000; 2000AU-0007489.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;

PI Nakata Y, Hasegawa M;

XX WPI; 2001-281978/29.

DR N-PSDB; AAD04201.

XX New biologically active complex comprising NR6 and

PT cardiotrophin-like-cytokine, for facilitating proliferation,

PT differentiation and/or survival of a cell -

XX Claim 32; Page 114-115; 123pp; English.

PS The present invention relates to a biologically active complex comprising
 CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).
 CC The complex is useful in the manufacture of a medicament for the
 CC treatment and/or prophylaxis of a subject, as it is involved in
 CC facilitating proliferation, differentiation and/or survival of a cell.
 CC The complex or its components have neurotrophic activity. The present
 CC sequence is human cardiotrophin-like cytokine (CLC) protein.

XX Sequence 223 AA;

Alignment Scores:

Pred. No.: 5,64e-92 Length: 223

Score: 1214.00 Matches: 223

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 83.26% Indels: 0
 DB: 22 Gaps: 0

US-09-931-704-1 (1-797) x AAE00828 (1-223)

QY 90 ATGACCTCCGAGCAGGGAGCTCGTGGGGATGTTAGCGTCTGTGACGCTGCTCTGG 149
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyThrValLeuTrp 20
 QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGGCCCTCCATC 209
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40
 QY 210 CAGAAACCTATGACCTACCGCTACCTGGAGCACCACCTCCGAGCTGGCTGGGACC 269
 Db 41 GlnLysThrThrAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGG 329
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProhrgLeuGly 80
 QY 330 CGAGAGACTGCCCAGGGCCACTGTGACTTGGAGGTGTGCGAAGGCTCAATGACAAA 389
 Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
 QY 390 CTGGCGCTGACCCAGACACTACAGGCGCTACAGCCACCTCTCTGTACTTGGTGGCCTC 449
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuGlyTyrLeuArgGlyLeu 120
 QY 450 AACCGTTCAGGCTGCCACTGCTGAGTGGCGCCGAGCTTGGCCACTTTCGACACGCTC 509
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 510 CAGGCGCTGCTGGGAGCATTCCGGCGCTCATGGAGCTCTGGGCTTACCCTGCCCCAG 569
 Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaLeuGlyTyrProLeuProGln 160
 QY 570 CGGCTGCTGGGAGTGAACCCACTTGGACTCTGGCTTGGCCCTCCAGTACGACTTCTCCAG 629
 Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
 QY 630 AAGATGGAGCACTTCTGGCTGTGAGGAGCTGCAGACTGCTGGCTGGCGCTCGGCCAAG 689
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 690 GACTTCAACCGCTCAAGAGAGATGCGACTTCCAGCTTCCAGAGCTGACGTCACCTGACCTG 749
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
 QY 750 GGGGCTCAT 758
 Db 221 GlyAlaHis 223

RESULT 12
 ID AAW29716
 AC AAW29716 standard; Protein; 225 AA.

XX AAW29716;

XX 09-NOV-1998 (first entry)

XX Mouse neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
 KW peripheral neuropathy; dystrophy; neural retina degeneration;
 KW common variable immunodeficiency; CVID; selective IgA deficiency;
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
 therapy.

Mus sp.
 OS
 XX Key
 FH Peptide
 FT
 FT Protein
 FT
 XX Location/Qualifiers
 PN 1..27
 XX /label= Sig_peptide
 FD 28..225
 XX /label= Mat_protein

XX WO9833922-A1.

XX 06-AUG-1998.

XX 02-FEB-1998; 98WO-US02363.

XX 30-JAN-1998; 98US-0016534.

XX 03-FEB-1997; 97US-0792019.

XX (AMGE-) AMGEN INC.

XX Chang M, Elliot GS, Sarmiento U, Senaldi G;

XX WPI; 1998-437475/37.

XX N-FSDB; AAV47512.

XX Newly isolated nucleic acid encoding human or murine neurotrophic factor NNT-1 - useful for treatment of neurological and immunological diseases or inflammation, also as vaccine adjuvant

XX Claim 13; Fig 5; 120pp; English.

XX This is the amino acid sequence of a murine neurotrophic factor, designated NNT-1, that is a growth factor for neurons and for B or T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512). Human NNT-1 (see AAW29715) is also provided. Vectors and host cells for use in the production of human murine recombinant NNT-1 polypeptides. These are used to treat: (i) neurological or immunological diseases, specifically Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and degeneration of the neural retina, or conditions characterised by T or B cell defects, e.g. common variable immunodeficiency (CVID), selective IgA deficiency, hypogammaglobulinaemia and X-linked agammaglobulinaemia (claimed), but many others disclosed; and (ii) inflammation. NNT-1 is also able to boost immunoreactivity and antibody production following vaccination, and, since it inhibits tumour necrosis factor production, it may also be useful for treating sepsis. In addition, cells that have been engineered to express NNT-1 can be implanted, or nucleic acids are delivered in gene therapy vectors.

XX Sequence 225 AA;

Alignment Scores:
 Pred. No.: 3,06e-90 Length: 225
 Score: 1193.00 Matches: 218
 Percent Similarity: 98.22% Conservative: 3
 Best Local Similarity: 96.89% Mismatches: 4
 Query Match: 81.82% Indels: 0
 DB: 19 Gaps: 0

US-09-931-704-1 (1-797) x AAW29716 (1-225)

QY 90 ATGACCTCCGAGCAGGGAGCTCGTGGGGATGTTAGCGTCTGTGACGCTGCTCTGG 149

Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyThrValLeuTrp 20

QY 150 CACCTCCCTGCGAGTCCAGCTCTCAATCCACAGGGGACCCAGGGCTGGCCCTCCATC 209

Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40

QY 210 CAGAAACCTATGACCTACCGCTACCTGGAGCACAACCTCCCGAGCTGGCTGGGACC 269

Db 41 GlnLysThrThrAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 270 TATCTGAACACTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCTGGGG 329
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 330 GCAGAGACTGCCCAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAA 389
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
 QY 390 CTGGGCTGACCCAGAACTACAGAGCCCTACAGCCACCTTCTGTGTACTTGGTGGCCCTC 449
 Db 101 LeuArgLeuThrGlnAsnTyrGlnAlaTyrSerHisLeuLeuGlyCysTyrLeuArgGlyLeu 120
 QY 450 AACCGTCAAGGCTGCCACTCTGAGCTGGCGCCGAGCCTGGCCCACTTCTGCACCCAGCTC 509
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 510 CAGGCGCTGCGGAGCATTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCGCAG 569
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 QY 570 CCGTCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCAGCTGCAGTCACTTCTCCAG 629
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
 QY 630 AAGATGAGCACTTCTGCTGTGAAGGAGCTGCAGACTGCTGGCTGTGGCTCGGCCAAG 689
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 690 GACTTCAACCCGCTCAAGAGAGATGCAGCTCCAGCAGCTGCAGTCACTTCTCCAGCTG 749
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 QY 750 GGGGCTCATGGCTTC 764
 Db 221 GluAlaHisGlyPhe 225

RESULT 13
 AAW56142
 ID AAW56142 standard; Protein; 225 AA.
 AC AAW56142;
 XX
 DT 13-JUL-1998 (first entry)
 XX
 DE Amino acid sequence of murine neurotrophic factor NNT-1.
 XX Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
 KW treatment; neurological disease; degeneration; Parkinson's disease;
 KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
 XX
 OS Mus sp.
 XX
 FH Key
 FT Peptide 1..27
 FT Protein /note= "signal peptide"
 FT /note= "mature peptide"
 XX
 PN US574172-A.
 XX
 PD 21-APR-1998.
 XX
 PF 03-FEB-1997; 97US-0792019.
 XX
 PR 03-FEB-1997; 97US-0792019.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Chang M;
 XX
 DR WPI; 1998-260526/23.
 DR N-PSDB; AAV22654.

XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
 PT useful for stimulating growth of motor and sympathetic neurons
 XX Claim 2; Fig 5; 41pp; English.
 XX The present sequence represents a murine neurotrophic factor, designated
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic
 CC neurons. The NNT-1 protein is useful in the treatment of neurological
 CC diseases characterised by the degeneration and death of particular
 CC classes of neurons. These diseases specifically include Parkinson's
 CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
 CC stroke and various degenerative disorders affecting vision.
 XX Sequence 225 AA;
 SQ

Alignment Scores:
 Pred. No.: 3,06e-90 Length: 225
 Score: 1193.00 Matches: 218
 Percent Similarity: 98.22% Conservative: 3
 Best Local Similarity: 96.89% Mismatches: 4
 Query Match: 81.82% Indels: 0
 DB: 19 Caps: 0

US-09-931-704-1 (1-797) x AAW56142 (1-225)

QY 90 ATGGACTCCGAGCAGGGGACTCGTGGGGATGTTAGCGTCCCTGTGCACGGTCTCTGG 149
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 150 CACCTCCCTGAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40
 QY 210 CAGAAACCTATGACTCCTCACCCTACCTGGAGCACCACCTCCGACCTGGCTGGGACC 269
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 270 TATCTGAACACTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGG 329
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 330 GCAGAGACTTGGCCAGGCGCACTGTGACTTGGAGGTGGCGAAGCCTCAATGACAAA 389
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
 QY 390 CTGGGCTGACCCAGAACTACAGAGGCTACAGCCACTTCTGTGTACTTGGTGGCCCTC 449
 Db 101 LeuArgLeuThrGlnAsnTyrGlnAlaTyrSerHisLeuLeuGlyCysTyrLeuArgGlyLeu 120
 QY 450 AACCGTCAAGGCTGCCACTCTGAGCTGGCGCCGAGCCTGGCCCACTTCTGCACCCAGCTC 509
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 510 CAGGCGCTGCGGAGCATTTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCCGCAG 569
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 QY 570 CCGTCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCAGCTGCAGTCACTTCTCCAG 629
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
 QY 630 AAGATGAGCACTTCTGCTGTGAAGGAGCTGCAGACTGCTGGCTGTGGCTCGGCCAAG 689
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 690 GACTTCAACCCGCTCAAGAGAGATGCAGCTCCAGCAGCTGCAGTCACTTCTCCAGCTG 749
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 QY 750 GGGGCTCATGGCTTC 764
 Db 221 GluAlaHisGlyPhe 225

RESULT 14
 ID AAY87814 standard; Protein; 225 AA.
 AC AAY87814;
 DT 24-AUG-2000 (first entry)
 DE Murine NNT-1 protein.
 KW NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment;
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
 KW retinopathy; immune disorder; hematopoietic disorder.
 OS Mus sp.
 PN US6054294-A.
 PD 25-APR-2000.
 PF 12-DEC-1997; 9TUS-0988819.
 PR 03-FEB-1997; 9TUS-0792019.
 PA (AMGE-) AMGEN INC.
 PI Chang M;
 DR WPI; 2000-338492/29.
 DR N-PSDB; AAA39483.
 PT New nucleic acids encoding neurotrophic factors useful for stimulating
 PT growth of motor or sympathetic neurons for treating neuron cell damage
 PS Claim 2b; Fig 5; 42pp; English.
 CC This invention describes a novel nucleic acid molecule (I) encoding a
 CC novel neurotrophic factor (NNT-1) (II) which has neurotropic,
 CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
 CC ophthalmological activity. (I) is useful for producing NNT-1
 CC polypeptides which are useful for treating patients in whom various
 CC cells of the central, autonomic, or peripheral nervous system have
 CC degenerated and/or have been damaged by congenital disease, trauma,
 CC mechanical damage, surgery, stroke, ischemia, infection, metabolic
 CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
 CC proteins are used to treat diseases like Alzheimer's, Parkinson's,
 CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
 CC disorders, and/or dystrophies or degeneration of the neural retina such
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
 CC night blindness, progressive cone-rod degeneration, immune disorders and
 CC hematopoietic disorders. (I) is effective in treating neurological
 CC conditions and promotes neuron regeneration. Neural functions are
 CC effectively restored in patients suffering from various neurological
 CC disorders. This sequence represents the murine NNT-1 protein described in
 CC the method of the invention.
 SQ Sequence 225 AA;
 Alignment Scores:
 Pred. No.: 3,06e-90 Length: 225
 Score: 1193.00 Matches: 218
 Percent Similarity: 98.22% Conservative: 3
 Best Local Similarity: 96.89% Mismatches: 4
 Query Match: 81.82% Indels: 0
 DB: 21 Gaps: 0
 US-09-931-704-1 (1-797) x AAY87814 (1-225)

QY 90 ATGGACCTCCGAGCAGGGGACTCGTGGGGGATGTTAGCGTCTGTGACCGTGTCTCTGG 149
 DB 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 150 CACCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209
 DB 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
 QY 210 CAGAAACCTATGACCTCACCGCTACCTGGAGACCACTCCGAGCTTGGCTGGGACC 269
 DB 41 GlnYsThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 270 TATCTGAATCTACTCGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGG 329
 DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 330 CGAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTGGGAGGCTCAATGACAA 389
 DB 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
 QY 390 CTGCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTTACTTGGTGGCTTC 449
 DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCystyrLeuArgGlyLeu 120
 QY 450 AACCGTCAAGGCTGCCACTGCTGAGCTGGCGCCGAGCTGGCCACTTCTGACACGACCTC 509
 DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 510 CAGGCGCTGCTGGGAGCATTGGCGGCTCATGGAGCTCTGGGCTACCCACTGCCGACG 559
 DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 QY 570 CGCTGCTGGGAGTGAACCCACTTGGACTCTGCTGGCCCTGCGCCACAGTCACTTCTCCAG 629
 DB 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
 QY 630 AAGATGGACGACTTCTGCTGCTGAAGGAGCTGCAGACCTGGCTGTGGCGCTCGGCAAG 689
 DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 690 GACTTCAACCGCTCAAGAGAGATGACGCTCCAGAGCTGCAGTCACTCCCTGACCTG 749
 DB 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 QY 750 GGGGCTCATGGCTTC 764
 DB 221 GluAlaHisGlyPhe 225
 RESULT 15
 ID AAY878177 standard; Protein; 225 AA.
 AC AAY878177;
 DT 05-JUN-2002 (first entry)
 DE Mouse novel neurotrophic factor NNT1.
 KW Mouse; NNT1; neurotrophic factor; IGE-related disease;
 KW Type I allergic disease; allergic rhinitis; eczema; dermatitis;
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
 KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;
 KW graft versus host disease; infertility; miscarriage; preterm labour.
 OS Mus sp.
 PN WO200215977-A2.
 PD 28-FEB-2002.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 16:20:23 ; Search time 5.7073 Seconds
(without alignments)
5635.701 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagtttcgagagcc.....tctctcttcgctccccccc 797

Scoring table:

BLOSUM62			
XGAPOP 10.0 , Xgapext 0.5			
YGAPOP 10.0 , Ygapext 0.5			
FGAPOP 6.0 , Fgapext 7.0			
DELTOP 6.0 , Delext 7.0			

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 244452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09931704/runat_27012003_154129_3760/app_query.fasta_1.7189
-DB=Published Applications AA -QPMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09931704 @CGN 1.1 51 @runat_27012003_154129_3760
-NCPV=6 -ICPV=3 -NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW PUB.pbp.*
2: /cgn2_6/ptodata/1/pubpaa/PTC_NEW PUB.pbp.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW PUB.pbp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pbp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW PUB.pbp.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pbp.*
7: /cgn2_6/ptodata/1/pubpaa/PTCUS_PUBCOMB.pbp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pbp.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW PUB.pbp.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pbp.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW PUB.pbp.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pbp.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW PUB.pbp.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pbp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1226	84.1	225	10	US-09-931-704-2
2	1193	81.8	225	10	US-09-931-704-5
3	885	60.7	164	10	US-09-864-761-40014
C 4	157.5	10.9	529	10	US-09-861-597-2

Alignment Scores:					
Pred. No.:	3.19e-80	Length:	225		
Score:	1226.00	Matches:	225		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	84.09%	Indels:	0		
DB:	10	Gaps:	0		

ALIGNMENTS

RESULT 1
US-09-931-704-2
; Sequence 2, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating Ige-Related Disease Using Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-931-704-2

5	151	10.4	27	10	US-09-864-761-40015	Sequence 40015, A
6	150	10.3	274	10	US-09-850-887-4	Sequence 4, Appli
7	141.5	9.7	507	9	US-10-078-547-24	Sequence 24, Appli
C 8	140	9.7	714	10	US-09-861-597-10	Sequence 10, Appli
9	139.5	9.6	503	9	US-10-078-547-2	Sequence 2, Appli
10	139	9.5	666	10	US-09-791-171-70	Sequence 70, Appli
11	138.5	9.5	802	10	US-09-823-240-2	Sequence 2, Appli
12	137.5	9.4	980	9	US-09-908-193-22	Sequence 22, Appli
13	137	9.4	1274	9	US-10-020-215-2	Sequence 2, Appli
14	136.5	9.4	280	10	US-09-835-232-1	Sequence 1, Appli
15	136.5	9.4	1567	9	US-09-835-232-2	Sequence 2, Appli
16	134	9.2	638	9	US-10-001-887-108	Sequence 108, App
17	131.5	9.0	865	9	US-09-957-998A-19	Sequence 19, Appli
18	129.5	8.9	527	9	US-09-854-133-216	Sequence 216, App
19	129.5	8.9	527	10	US-09-738-973-216	Sequence 216, App
20	128	8.8	674	10	US-09-925-299-979	Sequence 979, App
C 21	126.5	8.7	1004	9	US-09-738-626-5676	Sequence 5676, Ap
22	126	8.6	1247	9	US-09-908-193-2	Sequence 2, Appli
23	125.5	8.6	531	10	US-09-925-300-1444	Sequence 1444, Ap
24	125	8.6	1806	10	US-09-919-497-56	Sequence 56, Appli
25	124.5	8.5	171	10	US-09-864-761-34988	Sequence 34988, A
26	124.5	8.5	171	10	US-09-864-761-36844	Sequence 36844, A
C 27	124.5	8.6	606	10	US-09-861-597-4	Sequence 4, Appli
C 28	123.5	8.5	651	10	US-09-861-597-1	Sequence 1, Appli
29	123	8.4	2442	12	US-10-109-886-10	Sequence 10, Appli
30	122.5	8.4	1134	9	US-10-001-873-50	Sequence 50, Appli
C 31	122	8.4	606	10	US-09-861-597-6	Sequence 6, Appli
C 32	122	8.4	606	10	US-09-861-597-8	Sequence 8, Appli
33	121.5	8.3	847	9	US-10-143-133-2	Sequence 2, Appli
34	120.5	8.3	1422	10	US-09-735-933-1	Sequence 1, Appli
35	120.5	8.3	1690	10	US-09-788-043C-5	Sequence 5, Appli
36	120.5	8.3	4440	9	US-10-174-590-525	Sequence 525, App
37	120.5	8.3	4440	9	US-10-176-758-525	Sequence 525, App
38	120.5	8.3	4440	9	US-10-175-737-525	Sequence 525, App
39	120.5	8.3	4440	12	US-10-052-586-525	Sequence 525, App
40	120	8.2	726	10	US-09-770-689A-4	Sequence 4, Appli
41	120	8.2	803	10	US-09-770-689A-2	Sequence 2, Appli
42	120	8.2	881	10	US-09-816-860A-2	Sequence 2, Appli
43	120	8.2	1665	10	US-09-858-664A-2	Sequence 2, Appli
44	120	8.2	2630	9	US-10-077-130-2	Sequence 2, Appli
45	120	8.2	7968	9	US-10-077-130-5	Sequence 5, Appli

US-09-931-704-1 (1-797) x US-09-931-704-2 (1-225)

QY 90 ATGGACCTCCGACGAGGGGACTCGTGGGGGATGTAGCGTGTGTCACGGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGACGTGCGACGCTCAATCGCACAGGGGACCCAGGGCTGGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
QY 210 CAGAAACCTATGACCTCACCCGCTACTGGAGCACCACTCCGACGTTGGCTGGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGCTTCAACCTCCCGGCTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80
QY 330 GCAGAGACTCTGCCCGGGCCACTGTTGACTTGGAGGTGTGGCAAGCCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspLys 100
QY 390 CTGGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGTGGGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACGGTCAGGCTGCTGAGCTGGCGGCGAGCTGGCGCCACTTCTGACAGCCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGGCTGCTGGGACGACTTGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
QY 570 CCGCTGCTGGGAGTGAACCACTTGGAGCTCCTGGCCCTGCGCCACAGAGACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGAGCACTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCGCTGGCGCCAG 689
Db 181 LysMetAspAspPheTrpLeuLeuGlyGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGGCTCAAGAAAGATGCGACCTCCAGCAGCTGCGACCTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GlyAlaHisGlyPhe 225

RESULT 2

US-09-931-704-5
; Sequence 5, Application US/09931704
; Patent No. US20020041873A1

GENERAL INFORMATION:

; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using N
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931.704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5

LENGTH: 225

TYPE: PRT

ORGANISM: Murine

US-09-931-704-5

Alignment Scores:

Pred. No.: 7,14e-78 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 81.82% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-931-704-5 (1-225)

QY 90 ATGGACCTCCGACGAGGGGACTCGTGGGGGATGTAGCGTGTGTCACGGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGACGTGCGACGCTCTCAATCGCACAGGGGACCCAGGGCTGGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
QY 210 CAGAAACCTATGACCTCACCCGCTACTGGAGCACCACTCCGACGTTGGCTGGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGGCTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80
QY 330 GCAGAGACTCTGCCCGGGCCACTGTTGACTTGGAGGTGTGGCAAGCCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
QY 390 CTGGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGTGGGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCAGGCTGCTGAGCTGGCGGCGAGCTGGCGCCACTTCTGACAGCCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGGCTGCTGGGAGCATTGGGGGCTCATGCGAGCTCTGGGCTACCCACTGCCCCAG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
QY 570 CCGCTGCTGGGAGTGAACCACTTGGAGCTCCTGGCCCTGCGCCACAGAGACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AAGATGAGCACTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCGCTGGCGCCAG 689
Db 181 LysMetAspAspPheTrpLeuLeuGlyGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGGCTCAAGAAAGATGCGACCTCCAGCAGCTGCGACCTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GluAlaHisGlyPhe 225

RESULT 3

US-09-864-761-40014
; Sequence 40014, Application US/09864761
; Patent No. US20020048763A1

GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

```

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST HUMAN HIT: A1752561.1, EVALUE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 5,84e-56 Length: 164
Score: 885.00 Matches: 163
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 60.70% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-864-761-40014 (1-164)
QY 273 CTGAAGTACCTGGGCCCCCTTTCAACGAGCAGACTTCAACCCCTCCCGCTGGGGCA 332
Db 1 LeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20
QY 333 GAGACTCTCCCGAGGCCACTGTGTGACTTGGAGGTGTGGCGAGCCCTCAATGACAACTG 392
Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTyrArgSerLeuAsnAspLysLeu 40
QY 393 CGGTGACCCAGAACATACAGGCGCTACAGCCACTTCTGTGTTACTTGTGCGTGCCTCAAC 452

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST HUMAN HIT: A1752561.1, EVALUE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 5,84e-56 Length: 164
Score: 885.00 Matches: 163
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 60.70% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-861-597-2 (1-529)
QY 766 CAGAAGCCATGAGCCCGCCAGGTGCAGGGTGACTGCAGCTGGAGGCTGCATCTCTTC 707
Db 66 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaGly-SerGlyGlnGlnG 85
QY 706 TTGAGCCCGGTTCAAGTCCT-----TG 686
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlyProGlyGlyTyrGlyGlnGlnGlnG 105
QY 685 GCCGAGCGCCACAGCCAGGTCTTCAGAGTCTTCAGCAGCCAGGAAGTCGTCATCTTC 626
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlyProGlyGlyTyrGlyGlnGlnGlnG 105
QY 685 GCCGAGCGCCACAGCCAGGTCTTCAGAGTCTTCAGCAGCCAGGAAGTCGTCATCTTC 626
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlyProGlyGlyTyrGlyGlnGlnGlnG 105

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST HUMAN HIT: A1752561.1, EVALUE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 5,84e-56 Length: 164
Score: 885.00 Matches: 163
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 60.70% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-861-597-2 (1-529)
QY 766 CAGAAGCCATGAGCCCGCCAGGTGCAGGGTGACTGCAGCTGGAGGCTGCATCTCTTC 707
Db 66 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaGly-SerGlyGlnGlnG 85
QY 706 TTGAGCCCGGTTCAAGTCCT-----TG 686
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlyProGlyGlyTyrGlyGlnGlnGlnG 105
QY 685 GCCGAGCGCCACAGCCAGGTCTTCAGAGTCTTCAGCAGCCAGGAAGTCGTCATCTTC 626
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlyProGlyGlyTyrGlyGlnGlnGlnG 105
QY 685 GCCGAGCGCCACAGCCAGGTCTTCAGAGTCTTCAGCAGCCAGGAAGTCGTCATCTTC 626
Db 85 yProGlyGlyTyrGlyProArgGlnGlnGlyProGlyGlyTyrGlyGlnGlnGlnG 105
```

Db 105 yProSerGlyProGlySerAlaAlaAlaAlaSerAlaAlaAlaSerAlaGluSerGlyG1 125
Qy 625 AGGAGTCTACTGTGGCAGCGCCAGAG-----598
Db 125 yProGlyGlyTyTyGlyProGlyGlnGlnGlyProGlyGlyTyTyGlyProGlyGlnGlnG1 145
Qy 597 -----TCCAAGTGGTTCAGTCCAGGCGAGC 572
Db 145 yProGlyGlyTyTyGlyProGlyGlnGlnGlyProSerGlyProGlySerAlaAlaAlaAl 165
Qy 571 GGCTGGGCGAGTGGTAGCCAGAGTGCATGACGCCGCCCAATGTGCGCCAGCA-----517
Db 165 aAlaAlaAlaAlaSerGlyPro-----GlyGlnGlnGlyProGlyGlyTy 180
Qy 516 -----GGCCTGGAGCTGG-----TGCAGAGTGGGCC 488
Db 180 rGlyProGlyGlnGlnGlyProGlyGlyTyTyGlyProGlyGlnGlnGlyProSerGlyPr 200
Qy 487 AGGCTGGCGGCGAGCTCAG-----CAGTGGCAGGCTGACGG 452
Db 200 oGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaSerGlyProGlyGlnGlnGlyProGlyG1 220
Qy 451 TTGAGGCCAGCGCAAGTACACAGAGGTGCTGTAGGCT-----CCTAG 407
Db 220 yTyTyProGlyGlnGlnGlyProGlyGlyTyTyGlyProGlyGlnGlnGlyLeuSerG1 240
Qy 406 TTCTGGGTGAGCGCGAGTTGTCTATGAGGCTTCGCGCACACCTCCAGTCAACAGTGGCC 347
Db 240 yProGlySerAlaAlaAla-----AlaAlaAlaAlaGlyProGlyGlnGlnGlyPr 257
Qy 346 CTGGGCGAGAGTCTCTGCCCGGAGGGGGAGGGTGAAGTCTGGCTGTTGAAAGGGGG 287
Db 257 oGlyGlyTyTyGly-----ProGlyGlnGlnGlyProSer-----G1 269
Qy 286 CCAGGTAGTTCAGATGAGTCCAGCCAGCTGCGGAGTGGTGTCTCCAGTACGGGTG 227
Db 269 yProGlySerAla-----AlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyTyPr 288
Qy 226 AGGTATAGTCTTCTGGATGGAGGGGCCAGGCC--CTGGTCCCTGTGCGATTGAGA 170
Db 288 oGlyGlnGlnGlyProGlyGlyTyTyGlyProGlyGlnGlnGlyPro-----303
Qy 169 GCTGGCAGTCTCAGGAGTCTCCAGAGACCGCTGCACAGGCGACGCTAACATCCCCCAGAG 110
Db 304 -----SerGlyAlaGlySerAlaAlaAlaAlaAlaAlaAlaGlyProGlyG1 319
Qy 109 TCCCTGCTCGAGTCTCATGGGCTGG--GCTGGCGCGCGCGGTGCGGC-----60
Db 319 nGlnGlyLeuGlyGly-TyTyGlyProGlyGlnGlnGlyProGlyGlyTyTyGlyProGlyG 339
Qy 59 --TCCTCTCCGAGGCTGGGAGTGGAGGGGAGCGCGGCTCCGGCGAAGCT 6
Db 339 lnglnGlyProGlyGlyTyTyGlyProGlySerAlaSerAlaAlaAlaAlaAla 357

RESULT 5
US-09-864-761-40015
; Sequence 40015, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIORITY FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40015
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST_HUMAN HIT: A1457210.1, EVALUATE 4.60e-00
; OTHER INFORMATION: SWISSPROT HIT: P04258, EVALUATE 5.80e-01
US-09-864-761-40015

Alignment Scores:
Pred. No.: 0.000837 Length: 27
Score: 151.00 Matches: 26
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 96.30% Mismatches: 0
Query Match: 10.36% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-1 (1-797) x US-09-864-761-40015 (1-27)
Qy 275 GAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTTCCCGCTGGGGCAGA 334
Db 1 GluLeuProGlyProPheGlnArgAlaArgLeuGlnProSerProGlyGlyLys 20
Qy 335 GACTCTGCCAGGCGCACTGT 355
Db 21 AspSerAlaGlnGlyHisCys 27
RESULT 6
US-09-850-887-4
; Sequence 4, Application US/09850887
; Patent No. US20020009778A1

GENERAL INFORMATION:
APPLICANT: Lal, Preeti
Corley, Neil C.
Gorgone, Gina
TITLE OF INVENTION: THYROID AND PITUITARY MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/850,887
FILING DATE: 07-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: 09/087,678
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0535 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: g206712
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-850-887-4

Alignment Scores:
Pred. No.: 0.00136 Length: 274
Score: 150.00 Matches: 71
Percent Similarity: 34.56% Conservative: 4
Best Local Similarity: 32.72% Mismatches: 79
Query Match: 10.29% Indels: 63
DB: 10 Gaps: 14

US-09-931-704-1 (1-797) x US-09-850-887-4 (1-274)

Qy 15 GGAGCGCGGTCCCTCCCACT---CCGCCAGCTCCGGGA----- 54
Db 55 GlyProGlnArgProGlnGlyProPro-ProProGlyGlyProGlnGlnLysPr 74
Qy 55 -GAGGAGCGGACCGCGCGGCCAGCCAG-----CCCATGAGCTCGACAGGG 107
Db 74 oGlnGlyProPro-ProGlyGlyProGlnGlnArgProGlnGlyProProProG 94
Qy 108 GACTCGTGGGGATTTAGCGGTCTGTGACGGTCTGCGACCTCCCTGCACTGCA 167
Db 94 ly-----GlyProGlnGlnGlyProGlnGlyProProProGlyG 108
Qy 168 GCTCTCAATCCGAC-----AGGGGACCCAGGCGCTGGCCCCCTCC 206
Db 108 lyProGlnGlnArgProGlnGlyProProProProGlyGlyProGlnGlyProGlnG 128
Qy 207 ATCCAGAAAACCTATGACCTACCGCTACTGAGGACCACTCGGAGCTTGCTGGG 266
Db 128 lyPro-----ProProProGlyGly-----ProGlnGlnGlyProG 140

Qy 267 ACCTATCTGAACCTACCTGGGCCCCCTTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTG 326
Db 140 lNglYProProProGlyGlyProGlnGlnSerProProGlnGlyProProProProG 160
Qy 327 GGGGC---AGAGACTCTGCCAGGGCCACTGTTGACTTTGGAGGTGTGCGAAGCCTCAAT 383
Db 160 lyGlyProGlnGlnGlyProGlnGlyProProProProGlyGly-----ProGln- 176
Qy 384 CACAACTGCGGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTGTGTTACTTGGGT 443
Db 177 -----GlnSerProProGln-----G 182
Qy 444 GGCTCAACCGCTCAGGCTGCCACTGCTGAGCTGGCGCCGAGCTGGCCCACTTCTGCACC 503
Db 182 lyProProProProGly-----GlyProGlnGlnGlyPro-----GlnG 195
Qy 504 AGCTTCAGGCGCTGCTGGGAGCATTCGGGCGCTCATGGGAGCTCTGGGCTACCCACTG 563
Db 195 lyProProProGlyGlyProGlnGlnArgProGlnGlyArgGlnGly-----ProPro 211
Qy 564 CCCAGCGCTGCTGGGAGCTGAACCCACTTGGACTCCTGGCCCT 608
Db 211 roProGlyGly-ProGlnGlnAspProProGlnGlyProProPro 225

RESULT 7
US-10-078-547-24
Sequence 24, Application US/10078547
Publication No. US20020199211A1
GENERAL INFORMATION:
APPLICANT: Narayanaswamy Ramesh
APPLICANT: Miguel A. de la Fuente
APPLICANT: Ines M. Anton
APPLICANT: Raif S. Geha
TITLE OF INVENTION: WIP, A WASP-Associated Protein
FILE REFERENCE: 1242.1022-005
CURRENT APPLICATION NUMBER: US/10/078,547
CURRENT FILING DATE: 2002-02-19
PRIOR APPLICATION NUMBER: 09/599,287
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: PCT/US98/27501
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/101,457
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/068,533
PRIOR FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Fast-SEQ for Windows Version 4.0
SEQ ID NO 24
LENGTH: 507
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translated WIP ORF No. US20020199211A1 3
US-10-078-547-24

Alignment Scores:
Pred. No.: 0.00599 Length: 507
Score: 141.50 Matches: 83
Percent Similarity: 35.33% Conservative: 23
Best Local Similarity: 27.67% Mismatches: 97
Query Match: 9.71% Indels: 97
DB: 9 Gaps: 17

US-09-931-704-1 (1-797) x US-10-078-547-24 (1-507)

Qy 24 GCTCGCCCTCCCACTCCGCCAGCTCCGGAGAGGA----- 59
Db 127 SerArgProProLeuProGlyGlyArgSerThrSerAlaLysProPheSerPro 146
Qy 60 ---GCCGACCGCGCGG---CCGAGCCCGAGCCCTCCGAGCGGAGCTCG 113
Db 147 ProSerGlyProGlyArgPheProValProSerPro---GlyHisArgSerGly 163

```
QY 114 TGGGGGATGTTAGCGTGTGACGGTGTCTGGCACCTCCCTGCAGTCCAGCTCTC 173
Db 164 -----ProProGluProGlnArg 169
QY 174 AATCGCACAGGGACCCAGGCGCT-----GGCCCTCCATCCAGAAACCTATGACCTC 227
Db 170 AsnArgMetProProProArgProAspValGlySerLys--ProAspSerIle---Prop 188
QY 228 ACCCGTACTCGAGACACCACTCCGAGCTGGCTGGGACCTATCTGAACCTACTGGGC 287
Db 188 roProValProSerThrProArgPro---IleGlnSerSerLeuHisAsnArgGlySerP 207
QY 288 CCCCCTTTCAACGAGCAGACTTCAACCTCCCG-----CCTG 326
Db 207 roProValProGlyProArgGlnProSerProGlyProThrProProPheProG 227
QY 327 GGGCGAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAAGCCTCAATGAC 386
Db 227 lyAsnArgGlyThrAlaLeuGly-----GlyGlySerIleArgGlnSerProL 243
QY 387 AACTGGCGGTGACCCA---GAATACAGAGCCCTACAGCCACCTTCTGTGTTACTTGGCT 443
Db 243 euSerSerSerProPheSerAsnArgProProLeuProProThrProSerArgAlaL 263
QY 444 GG-----CCTCAACCGTCAGGC-----TGCCACTGCT 470
Db 263 euAspAspLysProProProProProProValGlyAsnArgProSerIleHisArgG 283
QY 471 GAGCTGGCGCGCAGCTGGCCCACTTCTGACACAGCCTCCAGGCGCTCTGGCGCAGCATT 530
Db 283 luAlaValProProProProGlnAsnAsnLysProProValPro-----298
QY 531 GCGGGCTCATGGCAGCTCTGGGTACCCACTGCGCCAGCGCTGCTGGGAGTCAACCC 590
Db 299 -----SerThrProArgProSerAlaProHisArg-----ProH 310
QY 591 ACTTG-----GACTCTCTGGCGCTGCCACAGTACTTCTCTCA 628
Db 310 isLeuArgProProProSerArgProGlyProProProLeuProProSerSerSerg 330
QY 629 -----GAAGATGGACGACTTCTGGCTGTGTGAAGAGAGTGCAGACCTGGTGTGGCGC 680
Db 330 lyAsnAspGluThrProArgLeu-----337
QY 681 TCGGCCAAGGACTTCAACCGCTCAAGAGAGATGCGACCTCCAGCAGCTGCAGTCACC 740
Db 338 -----ProGlnArgAsnLeuSerLeuSerSerSerThrProProLeuP 352
QY 741 CTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCGCTCCCGCC 796
Db 352 roSerProGlyArgSerGlyProLeuProProProValProSerGluArgProPro 370

RESULT 8
US-09-861-597-10
; Sequence 10, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 10
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-861-597-10

Alignment Scores:
Pred. No.: 0.00804 Length: 714
Score: 140.00 Matches: 70
Percent Similarity: 35.51% Conservative: 17
Best Local Similarity: 28.57% Mismatches: 101
Query Match: 9.68% Indels: 57
DB: 10 Gaps: 12

US-09-931-704-1 (1-797) x US-09-861-597-10 (1-714)
QY 690 CTTGGCCGAGCCGACAGCCAGGCTCTGAGCTCTTACAGCAGCCAGAGTCTGTCATCT 631
Db 43 ProGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyrGlyProGlyGln 62
QY 630 TCTGAGGAAGTCACTGTGGGAGCCAGGAGTCCAAGTGGTTTCAGTCCAGGACGCG 571
Db 63 GlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerGlyProGlySerAlaAla 82
QY 570 GCTGGGCGAGTGGGTAGCCGACAGCTGCCATGACGCGCCGCAATGCTGCCAGGAGCCCT 511
Db 83 AlaAlaAlaAlaAlaAla-----AlaGlyPro 91
QY 510 GGAGCTGTGTGAGAGTGGGCGAGGCTGGGCGAGCTCAGCAGTGGGAGCTGACGCTG 451
Db 92 GlyGlyTyr-----GlyProGly-----GlnGlnGlyProGlyGly 103
QY 450 TAGGCCAGCAAGTAACACAGAGGTGCTGTAGGCTCTGCTAGTCTCTGG-----400
Db 104 TyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerGly 123
QY 399 -----TCAGCCGAGTCTTGTCTATTGAGGCTTCGCCACACCT-----364
Db 124 ProGlySerAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyTyrGly 143
QY 363 CCAAGTCAACAGTGGCCCTGGGCGAGAGTCTCTCCCGGAGGAGGAGTGAAGTCTG 304
Db 144 ProGlyGlnGlnGlyProGlyGlyTyrGly-----ProGlyGlnGlnGlyProSer---160
QY 303 GCTCGTTGAAGGGGCGCCAGGTAGTTCAGATAGTTCGCCAGCCAGCTGCGGAGTTGGT 244
Db 161 -----GlyProGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGly 175
QY 243 GCTCCAGGTAGCGGTGAGTTCATAGGTTTCTGGATGGAGGCGCCAGGCGCTCGGTCCC 184
Db 176 -----GlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGly-----190
QY 183 CTGTGCGATTGAGAGCTGGCAGCTGCGAGGAGGTGCCAGACCGTGCACAGCAGCAGCTA 124
Db 191 -----GlnGlnGlyProSerGlyProGlySerAlaAlaAla 202
QY 123 ACATCCCCCAGAGTCCCTGCTCGGAGGTCCATGGGCGTGGG---GCTGGCGCGCGCG 67
Db 203 AlaAlaAlaAlaAlaAlaGlyProGlyGly-TyrGlyProGlyGlnGlnGlyProGlyGln 222
QY 66 GTGCGCG-----TCCTCTCCCGAGGCTGCGGAGTGGG---AGGCGAGCGCGCG 19
Db 222 yTyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerGln 242
QY 18 CTCCGGCGAGCT 6
Db 242 yProGlySerAla 246

RESULT 9
US-10-078-547-2
; Sequence 2, Application US/10078547
```



```
; Publication No. US20020199211A1
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Miguel A. de la Fuente
; APPLICANT: Ines M. Anton
; APPLICANT: Raif S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-005
; CURRENT APPLICATION NUMBER: US/10/078,547
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Human
; US-10-078-547-2

Alignment Scores:
Pred. No.: 0.00831 Length: 503
Score: 139.50 Matches: 80
Percent Similarity: 35.54% Conservative: 22
Best Local Similarity: 27.87% Mismatches: 88
Query Match: 9.57% Indels: 97
DB: 17 Gaps: 17

US-09-931-704-1 (1-797) x US-10-078-547-2 (1-503)
QY 24 GTCGCCCTCCCACTCCGCCAGCTCCGGGAGAGA----- 59
Db 124 SerArgProProLeuLeuProProGlyGlyArgSerThrSerAlaLysProPheSerPro 143
QY 60 ---GCGCACCCGGCGG---CCAGCCCCAGCCCATCGACTCCGACGAGGGACTCG 113
Db 144 ProSerGlyProGlyArgPheProValProSerPro---GlyHisArgSerGly----- 160
QY 114 TGGGGGATTTAGCTGCTGTGCACGCTGTCTGGCACCTCCCTCCAGTCCAGCTCTC 173
Db 161 -----ProProGluProGlnArg 166
QY 174 AATCGCACGGGACCCAGGGCCT-----GGCCCTCCATCCAGAAACCTATGACCTC 227
Db 167 AsnArgMetProProProArgProAspValGlySerLys--ProAspSerIle---Prop 185
QY 228 ACCCGCTACTGGAGCACCAACTCCGACGCTTGGCTGGGACCTATCTGAACCTACCTGGGC 287
Db 185 roProValProSerThrProArgPro---IleGlnSerSerLeuHisAsnArgGlySerP 204
QY 288 CCCCCTTTCAACAGCCAGACTTCAACCTCCCG-----CCTG 326
Db 204 roProValProGlyGlyProArgGlnProSerProGlyProThrProProPheProG 224
QY 327 GGGGAGAGACTCTCCAGGCGCACTGTGTGACTGGAGGTGGCGAGCCCTCAATGAC 386
Db 224 lyAsnArgGlyThrAlaLeuGly-----GlyGlySerIleArgGlnSerProL 240
QY 387 AAACCTGGCGGTGACCA---GAATACGAGGCGCTACAGCCACTTCTGTGTTACTTGGCT 443
Db 240 euSerSerSerProPheSerAsnArgProProLeuProProThrProSerArgAlaL 260
QY 444 GG-----CCTCAACCGTCAAGC-----TGCCACTGCT 470
Db 260 euAspAspLysProProProProProValGlyAsnArgProSerIleHisArgG 280
QY 471 GAGTGGCGCGGAGCTGGCCCACTTCTGCACCAAGCTCCAGGGCTCTGTGGGAGCATTT 530

; Sequence 70, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDLING, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-791-171-70

Alignment Scores:
Pred. No.: 0.00938 Length: 666
Score: 139.00 Matches: 76
Percent Similarity: 32.31% Conservative: 19
Best Local Similarity: 25.85% Mismatches: 99
Query Match: 9.53% Indels: 100
DB: 14 Gaps: 14

US-09-931-704-1 (1-797) x US-09-791-171-70 (1-666)
QY 18 GCGGGGCTGCCCTCCCACTCCGCCAGCTCCG-----GGAGAGAGAGCGCCACCG-- 69
Db 93 AlaAlaSerLysProProThrProProMet-ProIleAlaGlyProGluProAlaProPr 112
QY 70 -----GCCGGCCAGCCCGCCAGCC-----CCATG 92
```

```
Db 112 oLysProProThrProProMetProIleAlaGlyProGluProAlaProProLysProPr 132
Qy 93 GACCTCCGAGAGGAGGACTGCTGGGGAGTGTAGCGTGCCTGTGCAGCGTCTGGCAC 152
Db 132 oThr-ProProMetProIleAlaGlyProAlaProThrProGluSerGlnLeuAlaP 152
Qy 153 CTCCTGTCAGTCCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATCCAG 212
Db 152 roProArgProProThrProGlnThr-----ProThrGlyAlaProGlnGlnProG 169
Qy 213 AAAACCTATGACTCAC-----CGCTACCTCGGACCAAA 248
Db 169 luSerProAlaProHisValProSerHisGlyProHisGlnProArgArgThrAlaProA 189
Qy 249 CTCGCGAGCTTGGCTGGGACTATCGAACTACTGTGGGCCCCCTTTCAACAGAGCCAGA- 307
Db 189 laProPro-----TpAlaLysMetProIleGlyGluProProProAlaProSerArgp 207
Qy 308 -----CTTCAACCTCCCGCTGGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 341
Db 207 roSerAlaSerProAlaGluProProThrArgProAlaProGlnHisSerArgArgAlaA 227
Qy 342 CCAGGCGGCACTTGTACTTGGAGGTGGCGAAGCCTCAATGACAACTCGGCTGACC 401
Db 227 rgArgGlyHisArgTyArgThrAspThrGluArgAsnValGlyLysValAlaThrGlyP 247
Qy 402 CAGAACTACGA-----GGCC 416
Db 247 roSerIleGlnAlaArgLeuArgAlaGluAlaSerGlyAlaGlnLeuAlaProGlyT 267
Qy 417 TACAGCACCTTCTGTGTTACTTGGCTGGCTCAACGTCAGGCTCAGGCTGCCACTGTGAGCTG 476
Db 267 hrGluProSer-----ProAlaProLeuGly----- 275
Qy 477 CGCGGAGCGCTGGCCCACTTCTGCACGAGCTCCAGGCGCTGCTGGGCGAGCATTTGGCGGC 536
Db 276 -----GlnProArgSerTyLeuAlaProProThrArgProAla----- 288
Qy 537 GTCATGGCAGCTCTGGGCTACCCACTGCCCGCAGCGCTGCTGGGCTGAGTGAACCCACTGG 596
Db 289 -----ProThrGluProProProSer----- 295
Qy 597 ACTCCTGGCCCTGCCACAGTACTTCTCCAGAAGATGAGCAGCTTGGCTGCTGAAG 656
Db 296 -----ProSerProGln-----ArgAsnSerGlyArgArgAlaGluArgArgV 310
Qy 657 GAGCTGCAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAGAGATG 716
Db 310 alHisProAspLeuAla-----AlaGlnHisAlaAlaA 321
Qy 717 CAGCCTCCAGAGCTGAGTCAACCTGACCTGGGGG 754
Db 321 laGlnProAspSerIleThrAlaAlaThrThrGlyGly 333
```

RESULT 11

```
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: M0656/7064 (HCL)
; CURRENT APPLICATION NUMBER: US/09/823,240
; PRIOR FILING DATE: 2001-03-30
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Fa&#228;t-SEQ for Windows Version 3.0
```

```
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2
Alignment Scores:
Pred. No.: 0.0105 Length: 802
Score: 138.50 Matches: 71
Percent Similarity: 35.93% Conservative: 12
Best Local Similarity: 30.74% Mismatches: 84
Query Match: 9.50% Indels: 65
DB: 10 Gaps: 12
US-09-931-704-1 (1-797) x US-09-823-240-2 (1-802)
Qy 30 CCTCCCACTCCGCCA-----GCCCTCGGAGAGAGCGCCACCCGCGC---CGGCC 77
Db 381 ProProSerProProIleMetIleSerProProGlyLysAlaThrGlyProArgPro 400
Qy 78 AG-----CCCCAGCCCATGGACCTCCGAGC----- 103
Db 401 ValLeuProValCysValSerSerProValProGlnMetProProSerProThrAlaPro 420
Qy 104 AGGGGACTCGTGGGGATGTTAGCTGCTGTGTCACGGTGTCTGGCACCTCCCTGCAGT 163
Db 421 AsnGlySerLeuAspSerValThrTyProValSer-----ProProProThrSer 437
Qy 164 GCCAGCTCTCAATCGCACAGGGAGCCAGGGCTGGCCCTCCATCCAGAAAACCTATGA 223
Db 438 GlyProAlaAlaProProProProProProProProProProPro----- 453
Qy 224 CCTCACCGCTACCTGGAGCACCAACTCCGACGCTT-----GGCTGG 265
Db 454 ProProProLeuProProProProProProProProProLeuAlaSerLeuSerHisCysGlySer 473
Qy 266 GACCTATCTGAACCTACCTGGGCCCCCTTT-----CAACGAGCCAGACTT 310
Db 474 GlnAlaSerProProProGlyThrProLeuAlaSerThrProSerSerLysProSerVal 493
Qy 311 CAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTTGACTTGGAGGTGTG 370
Db 494 LeuProSerProSerAlaGlyAlaProAlaSerAla----- 505
Qy 371 GCGAAGCTCAATGACAACTCGCGTGCACCCAGAACTA----- 409
Db 506 -----GluThrProLeuAsnProGluLeuGlyAspSerSerAlaSerGlu 520
Qy 410 CGAGGCTACAGCCACTTCTGTGTACTTGGTGGCTCAACCGTCCAGGCTGCCACTGC 469
Db 521 ProGlyLeuGlnAlaAlaSerGlnProAlaGluSerProThrProGlnGlyLeuValLeu 540
Qy 470 TGAGCTGGCGCCAGCTGGCCACTTCTGCACCAGCTCCAGGCTGCTGGGCGCAT 529
Db 541 -----GlyProProAlaProProProProProProPro----- 554
Qy 530 TCGGGCGCTCATGGCAGCTCTG-----GGCTACCCACTGCCCGCCGCTGCC 577
Db 554 yProAlaTyAlaSerAlaLeuProProProProGlyProProProProProLeuPr 574
Qy 578 TGGGACTGAACCCACTTGGACTCTCGGCCCT 608
Db 574 oSerThrGlyProProProProProProPro 584
```

RESULT 12

```
US-09-908-193-22
; Sequence 22, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
```

Qy	527	-----CATTGCGGGCGTCA	TGCGAGCTCTGGGCTACCCA	560
Db	857	snGlnValGluAlaGluValIleValHisSerAspPheSerAlaSerAsnGlyAsnProA	877	
Qy	561	CTGCCCCAGCGCTGCTGGGACTGAACCCACACTGGACTCTCTGGCCCTGCCACACAGTCAC	620	
Db	877	sp-----LeuHisLeuGlnAspLeuGluProGluAspProL	889	
Qy	621	TTCCTCCAGAAGATGGACGACTTCTGGCT-----GCTGAAGGAGCTGCAG	665	
Db	889	euProProGluAlaProAspLeuIleSerGlyValGlyAspProGlyGlnGlyAlaAlaIat	909	
Qy	666	ACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACGGCTCAAGAA-----	709	
Db	909	rpLeuAspArgLeuGly--GlyCysGluLeuAlaAlaProGlyProAspArgLeuT	928	
Qy	710	-----GAAGATGCAGCTCCAGCGCTGCAGTCACCTC-----	742	
Db	928	hrCysLeuProGluAlaAlaSerAlaSerCysSerTyrProAspLeuGlnProGlyGluV	948	
Qy	743	-----GCACCTCGGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTTTCGCTC	791	
Db	948	alLeuGluGluThrProGlyAspSerCysGlnLeuIlysserProCysProLeuGlyAlaAs	968	
Qy	792	CCCCC	796	
Db	968	erPro	969	

[illegible][illegible]

QY	494	-----CITCTGCACACAGCTCCAGGGCCTCTCTGGGAG-----	526
Db	819	euthrArgAlaLeuLeu-----ProAlaGlyThrGlyGlnThrLeuLeuLeuGlnA	837
QY	526	-----	526
Db	837	laLeuValTyrAspAlaIleLysGlyAsnGlyArgLysLysSerProProAlaCysArgA	857

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:50:02 ; Search time 8.08533 Seconds
(without alignments)
5800.643 Million cell updates/sec

Title: US-09-931-704-1

Perfect score: 1458

Sequence: 1 attaaagctgcgcgagcc.....tctctcttctgcctccccccc 797

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO spo1/US09931704/runat 27012003 154127 3671/app query.fasta_1.7189
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=ral -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704 @CGN 1.1.89 @runat 27012003 154127 3671 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1226	84.1	225	1	US-08-792-019B-2
2	1226	84.1	225	3	US-09-106-182-2
3	1226	84.1	225	3	US-08-988-819-2
4	1226	84.1	225	4	US-09-016-534-2
5	1193	81.8	225	1	US-08-792-019B-5
6	1193	81.8	225	3	US-08-988-819-5
7	1193	81.8	225	4	US-09-016-534-5
8	160	11.0	1064	1	US-08-642-255-62
9	159	10.9	960	4	US-09-219-849-5
10	157.5	10.8	330	1	US-08-642-255-32
11	157.5	10.8	408	1	US-07-609-716-65
12	157.5	10.8	408	4	US-08-475-411A-65

13	157.5	10.8	408	4	US-08-478-029A-65	Sequence 65, Appl
14	157.5	10.9	529	4	US-09-247-806-2	Sequence 2, Appl
15	155.5	10.8	870	2	US-09-010-928B-2	Sequence 2, Appl
16	154	10.7	493	4	US-08-556-978B-59	Sequence 59, Appl
17	152	10.5	595	1	US-08-425-069-4	Sequence 4, Appl
18	152	10.5	595	2	US-08-317-844B-4	Sequence 4, Appl
19	148	10.2	252	1	US-08-642-255-61	Sequence 61, Appl
20	144.5	10.0	745	2	US-09-010-928B-28	Sequence 28, Appl
21	144	10.0	1008	4	US-09-219-849-8	Sequence 8, Appl
22	144	9.9	1057	3	US-08-931-820-4	Sequence 4, Appl
23	144	10.0	1065	1	US-08-642-255-80	Sequence 80, Appl
24	144	10.0	1065	3	US-08-642-246-16	Sequence 16, Appl
25	144	10.0	1065	4	US-09-451-208-16	Sequence 16, Appl
26	144	10.0	1065	5	PCT-US96-06229-16	Sequence 16, Appl
27	141	9.7	1341	3	US-08-963-825-18	Sequence 18, Appl
28	141	9.7	1341	4	US-09-500-811-18	Sequence 18, Appl
29	141	9.7	1341	4	US-09-570-573-18	Sequence 18, Appl
30	141	9.7	1341	4	US-09-548-608-18	Sequence 18, Appl
31	140	9.7	714	4	US-08-556-978B-61	Sequence 61, Appl
32	140	9.7	714	4	US-09-247-806-10	Sequence 10, Appl
33	139	9.5	905	2	US-08-574-959A-9	Sequence 9, Appl
34	139	9.5	905	4	US-09-357-014-9	Sequence 9, Appl
35	138.5	9.6	960	4	US-09-219-849-6	Sequence 6, Appl
36	138	9.5	1078	3	US-08-963-825-21	Sequence 21, Appl
37	138	9.5	1078	4	US-09-500-811-21	Sequence 21, Appl
38	138	9.5	1078	4	US-09-570-573-21	Sequence 21, Appl
39	138	9.5	1078	4	US-09-548-608-21	Sequence 21, Appl
40	137.5	9.4	1461	4	US-09-585-887-9	Sequence 9, Appl
41	137.5	9.4	1461	4	US-09-289-578-9	Sequence 9, Appl
42	137	9.4	357	1	US-07-609-716-66	Sequence 66, Appl
43	137	9.4	357	1	US-08-642-255-33	Sequence 33, Appl
44	137	9.4	357	4	US-08-475-411A-66	Sequence 66, Appl
45	137	9.4	357	4	US-08-478-029A-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-08-792-019B-2
; Sequence 2, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-792-019B-2

Alignment Scores: 3.75e-102 Length: 225
Pred. No.: 1226.00 Matches: 225
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 84.09% Gaps: 1
DB: 1

US-09-931-704-1 (1-797) x US-08-792-019B-2 (1-225)

QY 90 ATGGACCTCCGAGCGGGGACTCTGGGGGATGTTAGCTGCTGTGCACGGTGTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 150 CACTCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40

QY 210 CAGAAAACCTATGACCTACCGCTACCTGGAGCACCACTCCGAGCTTGGCTGGGACC 269
Db 41 GlnLysThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 270 TATCTGAACCTACCTGCGCCCTCTTCAACGAGCCAGACTTCAACCTCCCGCTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 330 GCAGAGACTCTGCCCCAGGGCCACTGTTGACTTGGAGTGTGGCGAGCTTCCACCCAGCCTC 389
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100

QY 390 CTGGCGCTGACCCAGAACCTACAGGCGCTACAGCCACTTCTGTGTTACTTGGCTGGGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGlnAlaTyrSerHisLeuLeuCysTyrProLeuProGln 160

QY 450 AACCGTCCGAGCTGCCACTGCTGAGCTGCCGCGAGCTTCCAGCAGCTGACCTGACCTG 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 510 CAGGGCTCTGCTGGGAGCATTGGGGCGTCATGCGAGCTTGGGCTACCCACTGCCCGCAG 569
Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaLeuGlyTyrProLeuProGln 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeu 180

QY 570 CCCTGCTGGGACTGAACCCACTTGGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 629
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 630 AAGATGACGACTTCTGGCTGCTGAAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 690 GACTTCAACCGGCTCAAGAGAGATGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220

QY 750 GGGGCTCATGCTTC 764
Db 221 GlyAlaHisGlyPhe 225

RESULT 2
US-09-106-182-2
; Sequence 2, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US

ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,182
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/051,053
FILING DATE: 30-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF385
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-182-2

Alignment Scores: 3.75e-102 Length: 225
Pred. No.: 1226.00 Matches: 225
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 3

US-09-931-704-1 (1-797) x US-09-106-182-2 (1-225)

QY 90 ATGGACCTCCGAGCGGGGACTCTGGGGGATGTTAGCTGCTGTGCACGGTGTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 150 CACTCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCCTGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40

QY 210 CAGAAAACCTATGACCTACCGCTACCTGGAGCACCACTCCGAGCTTGGCTGGGACC 269
Db 41 GlnLysThrTyAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 270 TATCTGAACCTACCTGCGCCCTCTTCAACGAGCCAGACTTCAACCTCCCGCTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 330 GCAGAGACTCTGCCCGAGGGCCACTGTTGACTTGGAGTGTGGCGAGCTTCCACCCAGCCTC 389
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100

QY 390 CTGGCGCTGACCCAGAACCTACAGGCGCTACAGCCACTTCTGTGTTACTTGGCTGGGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGlnAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 450 AACCGTCCGAGCTGCCACTGCTGAGCTGCCGCGAGCTTCCAGCAGCTGACCTTCCACCGCCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 510 CAGGGCTCTGCTGGGAGCATTGGGGCGTCATGCGAGCTTGGGCTACCCACTGCCCGCAG 569
Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaLeuGlyTyrProLeuProGln 160

QY 570 CCCTGCTGGGACTGAACCCACTTGGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeu 180

QY 630 AAGATGACGACTTCTGGCTGCTGAAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 690 GACTTCAACCGGCTCAAGAGAGATGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 749
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu 220

QY 630 AAGATGACGACTTCTGGCTGCTGAAGAGCTGCAGACCTGCTGGCTGGCTGGCCCAAG 689
| | | | |
Db 181 LysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
| | | | |
QY 690 GACTTCAACCGCTCAAGAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGCACCTG 749
| | | | |
Db 201 AspPheAsnArgLeuLysLysMetGlnProAlaAlaValThrLeuHisLeu 220
| | | | |
QY 750 GGGGCTCATGGCTTC 764
| | | | |
Db 221 GlyAlaHisGlyPhe 225
| | | | |
RESULT 3
US-08-988-819-2
; Sequence 2, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-988-819-2
Alignment Scores:
Pred. No.: 3,75e-102 Length: 225
Score: 1226.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 3 Gaps: 0
US-09-931-704-1 (1-797) x US-08-988-819-2 (1-225)
QY 90 ATGGACCTCCGAGGAGGAGCTGCTGGGGGATGTTAGCGTGCCTGTCAGCGTGTCTGG 149
| | | | |
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
| | | | |
QY 150 CACCTCCCTCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGGCTGGCCCTCCCATC 209
| | | | |
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
| | | | |
QY 210 CAGAAAACCTATGACTACCGCTACCTGGAGACCAACTCCGAGCTGGCTGGGACC 269
| | | | |
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
| | | | |

QY 270 TATCTGAACCTACTCTGGCGCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCTGGGG 329
| | | | |
Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80
| | | | |
QY 330 CGAGAGACTCTGGCCAGGGCCACTGTTGACTTGGAGGTGTGGCAGAGCTCAATGACAAA 389
| | | | |
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
| | | | |
QY 390 CTGGCGCTGACCCAGAACTACGAGGCTACAGCCACCTCTCTGTGTACTTGTGGTGGCCTC 449
| | | | |
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
| | | | |
QY 450 AACCGTCAAGGCTGCCACTGCTGAGCTGGCGCGCAGCCTGGCCCACTTTCACACGACCTC 509
| | | | |
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
| | | | |
QY 510 CAGGCGCTGCTGGGCGAGCATTCGCGGCGTCAATGGAGCTCTGGGCTACCCACTGCCCCAG 569
| | | | |
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
| | | | |
QY 570 CGGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGGCCCGCAGAGTCACTTCTCCAG 629
| | | | |
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
| | | | |
QY 630 AAGATGACGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGCTGGCTGGCGCTGGCCCAAG 689
| | | | |
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
| | | | |
QY 690 GACTTCAACCGCTCAAGAGAGATGACGCTCCAGCAGCTGCAGTCAACCTGCACCTG 749
| | | | |
Db 201 AspPheAsnArgLeuLysLysMetGlnProAlaAlaValThrLeuHisLeu 220
| | | | |
QY 750 GGGGCTCATGGCTTC 764
| | | | |
Db 221 GlyAlaHisGlyPhe 225
| | | | |
RESULT 4
US-09-016-534-2
; Sequence 2, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULLA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-016-534-2

Alignment Scores:

Pred. No.: 3,75e-102 Length: 225
Score: 1226.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.09% Indels: 0
DB: 4 Gaps: 0

US-09-931-704-1 (1-797) x US-09-016-534-2 (1-225)

QY 90 ATGGACCTCCGACGAGGGGACTCGTGGGGGATGTTAGCGTCCCTGTGCAGGTCCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCTGGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
QY 210 CAGAAACCTATGACTCACCCCGTACTGTGAGGACCACTCCGAGCTTGGCTGGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80
QY 330 GCAGAGACTCGCCAGGGCCACTGTTGACTGTGAGGTGTGGGAGGCTCAATGACAAA 389
Db 81 AlaGluThrLeuProAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
QY 390 CTGGCGTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGCTGGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCAAGCTGAGCTGAGTGGCGGAGCTGGCGGAGCTGGCCACTTCTGCAGCCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
QY 570 CCCTGCTGGGACTGAACCACTTGGACTCTGGCCCTGCCCAAGTACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AGATGGACGACTTCTGGCTGCTGAAGAGCTGAGACCTGGCTGGCTGGCTGGCCAG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGGCTCAAGAAAGATGATGAGCTCCAGCAGCTGCGAGTCAACCTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GlyAlaHisGlyPhe 225

RESULT 5

US-08-792-019B-5
; Sequence 5, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:

;
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC Compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792.019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-792-019B-5

Alignment Scores:

Pred. No.: 3,42e-99 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 81.82% Indels: 0
DB: 1 Gaps: 0

US-09-931-704-1 (1-797) x US-08-792-019B-5 (1-225)

QY 90 ATGGACCTCCGACGAGGGGACTCGTGGGGGATGTTAGCGTCCCTGTGCAGGTCCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCCTGGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40
QY 210 CAGAAACCTATGACTCACCCCGTACTGTGAGGACCACTCCGAGCTTGGCTGGGACC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACTGGGCCCCCTTTCAACGAGCAGACTTCAACCTCCCGCTGGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 330 GCAGAGACTTGGCCAGGGCCACTGTTGACTTGGAGGTGTGGCGAAGCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
QY 390 CTGGCGTGGACCCAGAACTACGAGGCTTACAGCCACTTCTGTGTACTTGGCTGGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCGTCAAGCTGAGCTGAGTGGCGGAGCTGGCGGAGCTGGCCACTTCTGCAGCCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGCGCTTGGGCGAGCATTCGGGGCGTTCATGGAGCTCTGGGCTACCCACTGCCCCAG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
QY 570 CCGTGTCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCCAAGTACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180

Qy	270	TATCTGAAC	TACCTGGG	CCCCCTTT	CAACAGC	AGCAGACTT	CAACCCCTCCCG	CCCTGGGG	320	
Db	61	TyrLeu	AsnTyrLeu	GlyPro	ProPhe	AsnGluPro	AspPhe	AsnProPro	ArgLeuGly 80	
Qy	330	GCAGAGACT	CTGCCAGG	GCACCTGTT	GACTTGAG	AGGTGGCG	AAGCCTCAAT	GACAA 389		
Db	81	AlaGlu	ThrLeuPro	ArgAla	ThrVal	AsnLeu	GluValTrp	ArgSerLeu	AsnAspArg 100	
Qy	390	CTGGCGG	CTGACCC	AGAACTAC	GAGGCTAC	GAGCCAC	CTTCTGTG	TTTACTT	GGTGGCCCTC 449	
Db	101	Leu	ArgLeu	ThrGln	AsnTyrGlu	AlaTyrSer	HisLeu	LeuCysTyrLeu	ArgGlyLeu 120	
Qy	450	AACCGT	CAGCTGCC	ACTGCTG	AGCTGCG	CCGAGCTG	CCGCCACTT	CTGCAC	CAAGCCTC 509	
Db	121	Asn	ArgGln	AlaAla	ThrAla	GluLeu	ArgSerLeu	AlaHisPhe	CysThrSerLeu 140	
Qy	510	CAGGGCT	GCTGGG	CAGCAT	TGGGGG	GTCA	TGGAGCTCT	GGGCTAC	CCCATGCCCCCAG 569	
Db	141	Gln	GlyLeu	LeuGlySer	IleAla	GlyValMet	AlaThrLeu	GlyTyrPro	LeuProGln 160	
Qy	570	CCGCTG	CTGGG	CTGAAC	CCACTT	GGACTT	CCCTGG	CCCTGCC	CACAGTGCATTCCTCCAG 629	
Db	161	Pro	LeuPro	GlyThrGlu	ProAla	TrpAlaPro	GlyPro	AlaHisSer	AspPheLeuGln 180	
Qy	630	AAGATG	GACGACTT	CTGGCTG	CTGAAG	GAGCTG	CAGACCT	GGCTGG	CGGCTCGGCCAAG 689	
Db	181	Lys	MetAsp	AspPhe	TrpLeu	LeuLysGlu	LeuGlnThrTrp	LeuTrp	ArgSerAlaLys 200	
Qy	690	GACTT	CAACCG	CTCAAGA	AGAGATG	CAGCTC	CAGAGCTG	CAGCTCA	CCCTGCACCTG 749	
Db	201	Asp	Phe	Asn	ArgLeu	LysIle	ValMet	GlnPro	ProAla	AlaSerValThrLeuHisLeu 220
Qy	750	GGGCT	CTCATG	CGCTC 764						
Db	221	Glu	Ala	His	GlyPhe 225					
RESULT 7										
US-09-016-534--5										
; Sequence 5, Application US/09016534										
; Patent No 6143874										
; GENERAL INFORMATION:										
; APPLICANT: CHANG, MING-SHI										
; APPLICANT: ELLIOTT, GARY S.										
; APPLICANT: SARMIENTO, ULLA										
; APPLICANT: SENALDI, GIORGIO										
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1										
; NUMBER OF SEQUENCES: 16										
; CORRESPONDENCE ADDRESS:										
; ADDRESSEE: AMGEN INC.										
; STREET: ONE AMGEN CENTER										
; CITY: THOUSAND OAKS										
; STATE: CA										
; COUNTRY: USA										
; ZIP: 91320										
; COMPUTER READABLE FORM:										
; MEDIUM TYPE: Floppy disk										
; COMPUTER: IBM PC compatible										
; OPERATING SYSTEM: PC-DOS/MS-DOS										
; SOFTWARE: PatentIn Release #1.0, Version #1.30										
; CURRENT APPLICATION DATA:										
; APPLICATION NUMBER: US/09/016,534										
; FILING DATE:										
; CLASSIFICATION:										
; PRIOR APPLICATION DATA:										
; APPLICATION NUMBER: US 08/792,019										
; FILING DATE: 03-FEB-1997										
; ATTORNEY/AGENT INFORMATION:										
; NAME: COOK, ROBERT R.										
; REGISTRATION NUMBER: 31,602										
; REFERENCE/DOCKET NUMBER: A-442B										
; INFORMATION FOR SEQ ID NO: 5:										
; SEQUENCE CHARACTERISTICS:										
; LENGTH: 225 amino acids										

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-016-534-5

Alignment Scores:
Pred. No.: 3,42e-99 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.88% Mismatches: 4
Query Match: 81.82% Indels: 0
DB: 4 Gaps: 0

US-09-931-704-1 (1-797) x US-09-016-534-5 (1-225)

```

QY 90 ATGGACCTCGAGCAGGAGCTGTGGGGATGTTAGCGTGTGCTGTCAGCGTCTCTGG 149
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
QY 150 CACCTCCCTGCGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGCTGGCCCTCCATC 209
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
QY 210 CAGAAACCTATGACCTACCCGCTACCTGAGCACCAACTCCGACCTGGCTGGAC 269
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 270 TATCTGAACCTACCTGGCCGCCCTTTCAACGAGCAGACTCAACCTCCCGCTGGG 329
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProArgLeuGly 80
QY 330 GCAGAGACTCTGCCAGGGGCACCTTTGACTTGGAGGTGGCGAAGCCTCAATGACAAA 389
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
QY 390 CTGCGGCTGACCCAGAACTACGAGGCTACAGCCACCTTTCTGTGTTACTTGGTGCTC 449
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 450 AACCTGTCAGCTGCTGAGCTGCGGCGGAGCTGCGCCACTTCTGACAGCCTC 509
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
QY 510 CAGGGCTGCTGGGACGATTGGGGCTCATGCAGCTCTGGGCTACCCACTGCCCCAG 569
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
QY 570 CCCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCTGCCCCACAGTGACTTCTCCAG 629
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
QY 630 AGATGACGACTTCTGGCTGTGAAGAGCTGAGACCTGGCTGGCTGGCTGGCCAG 689
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 690 GACTTCAACCGGCTCAGAGAAGATCAGCTCCAGCAGCTGAGTCACCTGCACCTG 749
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
QY 750 GGGGCTCATGGCTTC 764
Db 221 GluAlaHisGlyPhe 225

```

RESULT 8

US-08-642-255-62
Sequence 62, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A5556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 1064 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-62

Alignment Scores:
Pred. No.: 2,61e-06 Length: 1064
Score: 160.00 Matches: 80
Percent Similarity: 37.66% Conservative: 7
Best Local Similarity: 34.63% Mismatches: 81
Query Match: 10.97% Indels: 63
DB: 1 Gaps: 15

US-09-931-704-1 (1-797) x US-08-642-255-62 (1-1064)

```

QY 4 AAAGTCTCGCGAGCCGCGGCTCGCCCTCCACTCCGACGCTCCGGAGAG----- 57
Db 135 ArgGlyAspProGlyProGlyAlaProGlyProAlaGlyProGlySerArgGly 154
QY 58 GAGCGCACCGCGCGCCGACCCAGCCGCGCCATGAGACTCCG---AGCAGGGACTCGT 114
Db 155 AspProGlyPro-ProGlyAlaProGlyProAlaGlyProGlySerArgGlyAspPr 174
QY 115 GGGGATGTTAGCGTGTGTCAGCGTCTCTGGCACCTCCCTGCGAGTCCAGCTCTCA 174
Db 174 oGlyProProGlyAlaPro-----GlyProAlaGlyProPro-----G1 187
QY 175 ATCGCACAGGGAGCCAGGGCT-----GGCCCTCCATCCAGA 213
Db 187 ySerArg-GlyAspProGlyProGlyAlaProGlyProAlaGly-ProProGlySer 206
QY 214 AAACCTATGACTCACCCTACCTGGAGCAGCAACTCCGAGCTGGCTGGGACCTATC 273
Db 207 ArgGlyAsp-ProGlyPro---ProGlyAlaProGlyPro----- 218
QY 274 TGAACCTACTGGGCGCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGG----- 328
Db 219 -----AlaGlyProProGlySerArgGly-----AspProGlyProProGlyAlaPr 234
QY 329 -----GGCAGAGACTCTGCCCGAGGCCACTGTTGACTTGGAGGTGTGGCGAAGCC 378
Db 234 oGlyProAlaGlyProProGlySerArgGlyAspProGlyProGlyAlaProGlyPr 254
QY 379 TCAATGACAAACTGGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACT 439

```

Db 254 o-----AlaGlyProProGlySerArgGly----- 262
 QY 439 TGGCGTGGCTCAACCTCAGGCTGCCACTGCTGAGCTGGCGCCGAG- 484
 Db 263 ----AspProGlyProProGlyAlaHisGlyProAlaGlyProGlySerGlyAlaHisGlyPr 291
 QY 485 -CCTGCGCCACTT---CTGCACGAGCTCCAGGGCTCTGGGGCAGCATTTGGCGGGCGTCA 540
 Db 281 oAlaGlyProGlySerGlyAlaHisGlyProAlaGlyProGlySerGlyAlaHis--GlyProAl 300
 QY 541 TGGCAGCTCTGGGCTACCACTCCCGAGCGCTGCTGGGACTGAA-----C 588
 Db 300 aGlyProGlySerGlyAlaProGly--ProAlaGlyProProGlySerArgGlyAspProGlyP 320
 QY 589 CCACTTGGACTCCTGGCCCTGCC 611
 Db 320 roProGlyAlaProGlyProAla 327
 RESULT 9
 US-09-219-849-5
 ; Sequence 5, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHIELE D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; TYPE: PRT
 ; LENGTH: 960
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-5
 Alignment Scores:
 Pred. No.: 3,1e-06 Length: 960
 Score: 159.00 Matches: 74
 Percent Similarity: 40.82% Conservative: 6
 Best Local Similarity: 37.76% Mismatches: 87
 Query Match: 10.91% Indels: 29
 DB: 4 Gaps: 11
 US-09-931-704-1 (1-797) x US-09-219-849-5 (1-960)
 QY 46 CCTCCGGGA---GAGGAGCGCACCAGCGCGCCGCCAGCCAGCCCATGAGCTCCGAG 102
 Db 111 ProGlySerArgAspProGlyPro-ProGlyAlaProGlyProAlaGlyProProGly 130
 QY 103 CAGGGGACTCGTGGGGCTTGTACGCTGTGACGGTGTCTGGCAGCTCCCTCAG 162
 Db 130 ySerArgAspProGlyProProGlyAlaPro-----GlyProAlaGlyProProGlySe 148
 QY 163 TGGCAGCTCTCAATCGCACAGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATG 222
 Db 148 rArgAspPro-GlyProProGlyAlaProGlyPro-AlaGlyPro-ProGlySerArgAs 167
 QY 223 ACTCACCCCTACTGAGACCACTCCGAGCTTGGCTGGGACCTATCTGAACTACC 282
 Db 167 pProGlyPro---ProGlyAlaProGlyPro---AlaGlyProProGlySerArgAsPr 185

QY 283 TGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGC 342
 Db 185 OGlyProPro-----GlyAlaProGlyProAlaGlyProProGlySerArgAsp----- 201
 QY 343 CAGGGCCACTGTTGACTTGGAGGTGTGGAGAGCTCAATGACAACTGCGGCTGACCC 402
 Db 202 -ProGlyProProGlyAlaProGlyProAlaGlyPro-----ProGlySerArgAspPr 219
 QY 403 AGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGTGGCCTCAACCGTCAGGCTG 462
 Db 219 OGlyProProGlyAlaHisGlyProAla-----GlyProGlyAla 233
 QY 463 CCACTCTGAGCTGCGCCGCGAGCTGCGCCACTTCTGCACGAGCTCCAGGGCTGTGG 522
 Db 233 aHisGlyProAlaGlyProGlyAla-----HisGlyProAlaGlyProGlySGL 250
 QY 523 GCAGCATTTGGGGCTGATGGCAGCTCTGGGTACCCAGTGGCCCGCCGCTGCTGGGA 582
 Db 250 yAlaHis--GlyProAlaGlyProGlyAlaProGly--ProAlaGlyProProGlyS 269
 QY 583 CTGAA-----CCCACTTGGACTCTGGCCCTGCC 611
 Db 269 erArgAspProGlyProProGlyAlaProGlyProAla 281
 RESULT 10
 US-08-642-255-32
 ; Sequence 32, Application US/08642255
 ; Patent No. 5773249
 ; GENERAL INFORMATION:
 ; APPLICANT: CAPPELLO, Joseph
 ; APPLICANT: FERRARI, Franco A.
 ; TITLE OF INVENTION: High Molecular Weight Collagen-Like
 ; TITLE OF INVENTION: Protein Polymers
 ; NUMBER OF SEQUENCES: 135
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/642,255
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROWLAND, Bertram I.
 ; REGISTRATION NUMBER: 20,015
 ; REFERENCE/DOCKET NUMBER: A55556-3/BIR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 494-8700
 ; TELEFAX: (415) 494-8771
 ; TELEX: 910 277299 FHT UR
 ; INFORMATION FOR SEQ ID NO: 32:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 330 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-642-255-32
 Alignment Scores:
 Pred. No.: 2,96e-06 Length: 330
 Score: 157.50 Matches: 72
 Percent Similarity: 37.00% Conservative: 2
 Best Local Similarity: 36.00% Mismatches: 90

NUMER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,411A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-411A-65

Alignment Scores:
Pred. No.: 3,178-06 Length: 408
Score: 157.50 Matches: 72
Percent Similarity: 37.00% Conservative: 2
Best Local Similarity: 36.00% Mismatches: 90
Query Match: 10.80% Indels: 36
DB: 4 Gaps: 12

US-09-931-704-1 (1-797) x US-08-475-411A-65 (1-408)
QY 30 CCTCCACTCCGACGCTCCGGAGAGGAGCGCACCGCGCGCCAGCCGCCACCC 89
Db 41 ProProGlyProProGlyProProGlyProProGlyProProGlyProProGlyPro 60
QY 90 ATGGACCTCGAGC---AGGGACTCGTGGGGATTTAGCGTCT---GTGCAGGTG 143
Db 60 roGlyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyProProGlyP 80
QY 144 CTCTGGCACTCCCTCGAGTCCAGTCTCAATCGCACAGGGAGCCAGGGCGCTGGCCCC 203
Db 80 roProGlyProProGlyProProGlyProProGlyProProGlyProProGlyProProG 100
QY 204 TCCATCCAGAAACCTATGACCTCACCGCTACCTGGAGCACCACCACTCCGACGCTTGGCT 263
Db 100 lyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyPro----- 116
QY 264 GGGACCTATCTGAACCTACTGGGCGCCCGCTTTCAACGAGCGAGACTTCAACCTCCCGC 323

Db 117 -----ProGlyProProGly-----ProProGlyProP 126
QY 324 CTGGG-----GGCAGAGACTCTGCCAGGCGCACTGTGACTTGGAGGTGGCGAAGC 377
Db 126 roGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyProAlaGlyP 146
QY 378 CTCAATGACAAACTCGGCTGACCCAGAACACTACGAGGCTACAG---CCACTTCTGTGT 434
Db 146 ro-----ValGlySerProGlyAlaProGlyProProGlyProProGlyPro- 161
QY 435 TACTTGGCTGGCTCAACCGTCAGCTGCCACTGCTGAGCTGGCGGCGAGCTGGCCCCAC 494
Db 162 -----ProGlyProProGlyAlaProGlyProProGlyPro- 176
QY 495 TTCTGCACCGAGCTCCAGGCGCTGTGGCGAGCATTTGGCGGCTCATGCGAGCTCTGGGC 554
Db 177 -----ProGlyProProGlyProAlaGlyPro- 194
QY 555 TACCCACTGCCCGCGCTGCTGGGACTGAACCCACTTGGAGCTCTGGCCT 608
Db 194 roProGlyProProGlyPro-ProGly-----ProProGlyAlaProGlyPro 209
RESULT 13
US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

! MOLECULE TYPE: protein
US-08-478-029A-65

Alignment Scores:
Pred. No.: 3.17e-06 Length: 408
Score: 157.50 Matches: 72
Percent Similarity: 37.00% Conservatives: 2
Best Local Similarity: 36.00% Mismatches: 90
Query Match: 10.80% Indels: 36
DB: 4 Gaps: 12

US-09-931-704-1 (1-797) x US-08-478-029A-65 (1-408)

```
QY 30 CTTCCACTCCGAGCTCGGAGAGGAGCCACCCCGCCGCGCCAGCCGCCCC 89
Db 41 ProProGlyProProGly-ProProGlyAlaProGlyProProGlyPro-Prop 60
QY 90 ATGGACTCCGAGC--AGGGACTCTGGGGGATTTAGCGTCCT---GTGCACGGTG 143
Db 60 roGlyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyProGlyP 80
QY 144 CTCTGGCACTCTCCAGTCAGTCAGCTCTCAATCGACAGGGGACCCAGGCCCTGGCCCC 203
Db 80 roProGlyProProGlyProProGlyAlaProGlyProProGlyProProG 100
QY 204 TCCATCCAGAAACCTATGACCTCACCCGCTACCTGGGACCACTCCGCGACTTGGCT 263
Db 100 lyProProGlyProAlaGlyProValGlySerProGlyAlaProGlyPro- 116
QY 264 GGGACCTATCTGAACCTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCTCCCGCC 323
Db 117 -----ProGlyProProGly-----ProProGlyProp 126
QY 324 CTGGG-----GGCAGAGACTCTCCAGGCGCCACTGTTGACTGGAGGTGGGAGC 377
Db 126 roGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyProAlaGlyP 146
QY 378 CTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTACAG--CCACTTCTGTGT 434
Db 146 ro-----ValGlySerProGlyAlaProGlyProProGlyProProGlyPro- 161
QY 435 TACTTGGCTGCTCAACCGTCAAGCTGCCACTGCTGAGCTGCGCCGCGAGCTGGCCAC 494
Db 162 -----ProGlyProProGlyAlaProGlyProProGlyPro--ProGlyPro- 176
QY 495 TTCTGACCACTCTCCAGGCTCTGGGAGCATTGGGGGCTCATGGCAGCTCTGGCC 554
Db 177 -----ProGlyProProGlyProAlaGlyPro---ValGlySerProGlyAlaProGlyP 194
QY 555 TACCCTGCTCCCGCCGCTGCTGGGACTGAACCCACTTGGACTCTCTGGCCCT 608
Db 194 roProGlyProProGlyPro-ProGly-----ProProGlyAlaProGlyPro 209
```

RESULT 14

US-09-247-806-2
; Sequence 2, Application US/09247806
; Patent No. 6280747
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/247,806
; CURRENT FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: FR 98/01614
; EARLIER FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529

! TYPE: PRT
! ORGANISM: Nephila clavipes
US-09-247-806-2

Alignment Scores:
Pred. NO.: 3.46e-06 Length: 529
Score: 157.50 Matches: 91
Percent Similarity: 32.81% Conservatives: 14
Best Local Similarity: 28.44% Mismatches: 122
Query Match: 10.89% Indels: 93
DB: 4 Gaps: 15

US-09-931-704-1 (1-797) x US-09-247-806-2 (1-529)

```
QY 766 CAGAGCATGAGCCCCCAGGTGTCAGGTGAGTGTGAGGCTGAGGCTGCTCTTCTTC 707
Db 66 GlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaGly-SerGlyGlnGlnG 85
QY 706 TTGAGCCGTTGAAGTCTCT-----TG 686
Db 85 yProGlyGlyTyrglyProArgGlnGlnGlyProGlyGlyTyrglyGlnGlnGlnG 105
QY 685 GCCAGAGCCGACAGCCAGGTCTGAGCTCTTCAGCAGCAGAGTCGTCCATCTTCTGG 626
Db 105 yProSerGlyProGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyG 125
QY 625 AGGAAGTCACTGTGGCGGCGCAGGAG----- 598
Db 125 yProGlyGlyTyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnG 145
QY 597 -----TCCAAAGTGGGTTTCAGTCCCGGCGAGC 572
Db 145 yProGlyGlyTyrglyProGlyGlnGlnGlyProSerGlyProGlySerAlaAlaAla 165
QY 571 GGCTGGGCGAGTGGGTAGCCAGAGCTGCATGACGCCCGCATGTGTCGCCAGCA----- 517
Db 165 aAlaAlaAlaAlaSerGlyPro-----GlyGlnGlnGlyProGlyGlyTy 180
QY 516 -----GGCCCTGGAGGCTGG-----TGCAGAAAGTGGGCC 488
Db 180 rGlyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyProSerGlyP 200
QY 487 AGGCTCGCGCGCAGCTCAG-----CAGTGGCAGCTGACGG 452
Db 200 oglySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyP 220
QY 451 TTGAGCCGACCGAAGTAACACAGAGGTGGCTGAGGCT-----CGTAG 407
Db 220 yTyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyLeuSerG 240
QY 406 TTCTGGGTGAGCGCGAGTTTGTTCATTGAGGCTTCGCCACACCTCCCAAGTCAACAGTGGCC 347
Db 240 yProGlySerAlaAlaAla-----AlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyP 257
QY 346 CTGGGCGAGAGTCTCTCCCGCCAGCGGGAGGTTGAAGTCTGGCTGTTGAAGGGGGG 287
Db 257 oglyGlyTyrgly-----ProGlyGlnGlnGlyProSer-----G 269
QY 286 CCCAGTAGTTCAGATAGTCCCGCAGCAAGCTCGGAGTTGTTGCTCCAGTAGCGGGTG 227
Db 269 yProGlySerAla-----AlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyrglyP 288
QY 226 AGGTTCATAGTTTCTGGATGGAGGGCGCAGGCC-----CTGGGTCCTGTGCGATTGAGA 170
Db 288 oglyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyPro----- 303
QY 169 GTTGCACCTGCGAGGAGTGCAGACACCGTGCACAGGACCGCTAACATCCCCACGAG 110
Db 304 -----SerGlyAlaGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlyG 319
QY 109 TCCCTGTCTCGAGGTCCATGGGCTGGG---GCTGGGCGCGCGGTGGCGG----- 60
Db 319 nGlnGlyLeuGlyGly-TyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyG 339
```

QY 59 --TCCTCTCCGGAGCTGGCGAGTGGAGGCGGAGCGCGGCTCCGGCGAAGCT 6
Db 339 lngInGlyProGlyGlyTyrGlyProGlySerAlaSerAlaAlaAlaAla 357

RESULT 15
US-09-010-928B-2
; Sequence 2, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-2

Alignment Scores:

Pred. No.:	6.18e-06	Length:	870
Score:	155.50	Matches:	76
Percent Similarity:	40.17%	Conservative:	20
Best Local Similarity:	31.80%	Mismatches:	108
Query Match:	10.75%	Indels:	35
DB:	2	Gaps:	10

US-09-931-704-1 (1-797) x US-09-010-928B-2 (1-870)

QY 701 CCGGTTGAAGTCTTGGCGGAGCGCCACAGCCAGGCTCCAGCTCTCCAGCAGCCAGAA 642
Db 58 ProAnGluGluPheValHisGluValGlnAspLeuIleGlnMetLeuSerGlnGluGln 77
QY 641 GTCGTCCATCTTCTGGAGGAAGTCACTGTGGGAGGAGCCAGGA-----GTCCAAAGTGGG 588
Db 78 Ile-----AenGluValAspThr-SerGlyProGlyGlnTyrTyrArgSerSe 93
QY 587 TTCAGTCCAGGCGGCTGGGGTAGGTGGTACCCAGAGCTGCCANTGACCCCGCAAT 528
Db 93 rSerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 113
QY 527 GCTGCCAGCAGGCGCCCTGGAGGCTGGTGAGAGTGGGCGGAGGCTGGCGGCGAGCTCAGC 468
Db 113 i-ThrValGlyGlySerGlyGlyGlyGlyGlnProSerGlyAlaGlyProSerGlyThrGlyG 133
QY 467 AGTGGCAGCCTGAGGCTTGAAGCCAGCCAGTAACACAGAGGTTGGCTGTAGGCTCGTA 408
Db 133 lyTyAlaProThrGlyTyrAlaProSerGlySerGlyAlaGlyGlyValArgProSerA 153

Search completed: January 27, 2003, 16:25:43
Job time : 14.0853 secs

QY 407 GTTCTGGTCCAGCCGACAGTTTGTTCATTAGAGCTTCGCCACACCTCCCAAGTCAACAGTGGC 348
Db 153 laSerGlyProSerGlySerGlyProSerGlyGlySerArgProSerSerSerGlyProSer 173
QY 347 CTTGGGAGAGTCTCTGCCCCCAGCGGGAGGTTGAAGTCTGGCTCGTTGAAAGGGGG 288
Db 173 erGlyThrArgProSerProAenGlyAlaSerGlySerSer-----ProGlyG 189
QY 287 GC-----CCAGGTAGTTTCAGATAGTCCAGCCAGCAAGCTGGCGAGTGTGTCTCCAGGTA 234
Db 189 lyIleAlaProGlyGlySerAen-----SerGlyGlyAlaGlyValSerGlyA 205
QY 233 CGGGTGAAGTCATAGGTTTCTGATGAGAGGGGCCAGGCCCTCGCTCCCTGTGGATT 174
Db 205 laThrGlyGlyProAlaSerSerGly-----SerTyrGlyPro----- 217
QY 173 GAGAGCTGGCACTGCAGGAGGTGCCAGACACCGTGCACAGGCACGCTTAACATCCCCCA 114
Db 218 -----GlySerThrGlyGlyThrTyrGlyProSerGlyGly-----SerGluProp 233
QY 113 CGAGTCCC---CTGCTCGAGGTCCATGG-----GGCTGGGGCTGGGCGCGCGGCTCCG 63
Db 233 heGlyProGlyValAlaGlyGlyProTyrSerProGly-GlyAlaGlyProGlyGlyAla 252
QY 62 GGCTCCTCTCCCGAGGCTGGCGGAGTGGAGGGCGAGCGCGCGGCTCCG 14
Db 253 GlyGlyAlaTyrGly-ProGlyGlyValGlyThrGlyGlyAlaGlyPro 268

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:49:37 ; Search time 150.265 Seconds
(without alignments)
6508.980 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aactcgagtgaggcctggc.....ccttgctaagtctctca 5087

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPRO_spool/US09931704/runat_27012003_154126_3648/app_query.fasta_1.7189
-DB=PIR_73 -QMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09931704 @CGN 1 1 351 @runat_27012003_154126_3648 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	382.5	4.1	1585	2	hypothetical prote
2	339	3.6	1492	2	collagen alpha 1' (
3	336	3.6	1188	2	extensin-like prot
4	336	3.6	2715	2	eyelid - fruit fly
5	333	3.6	1414	1	collagen alpha 2(I
6	331.5	3.6	1486	1	collagen alpha 1(I
7	329.5	3.5	3570	2	mucin MUC5B, trach
8	327.5	3.5	1466	1	collagen alpha 1(I
9	325	3.5	1464	2	collagen alpha 1(I
10	323	3.4	2187	2	nascent polypeptid
11	320	3.4	1691	1	collagen alpha 5(I
12	319.5	3.4	1006	2	hypothetical prote
13	314.5	3.3	620	2	hydroxyproline-ric
14	313.5	3.3	2142	2	MHC class III hist

15	313	3.3	1184	2	S50832
C 16	311	3.3	1487	1	CGHU6C
17	310	3.3	1690	1	CGHU1B
18	309.5	3.3	1344	1	A35175
C 19	307	3.3	1497	2	I49607
C 20	306.5	3.3	1453	2	S21826
C 21	305	3.3	1418	2	T45467
C 22	304.5	3.3	1049	1	CGB07S
C 23	301	3.2	4957	2	T03455
24	299	3.2	1870	2	S37671
25	297.5	3.2	1670	1	CGHU3B
26	297.5	3.2	1872	2	S36152
27	296.5	3.1	1184	2	G01763
C 28	296.5	3.2	1366	1	CGHU2S
C 29	296.5	3.2	1690	1	CGHU1B
C 30	295.5	3.2	1106	2	J04005
C 31	294.5	3.2	2944	2	A54849
C 32	293	3.1	1752	2	A45407
C 33	292.5	3.1	1496	1	CGHU2V
C 34	292	3.1	886	2	I50694
C 35	291.5	3.1	1487	2	B41182
C 36	290	3.1	1042	1	CGCH1S
C 37	290	3.1	1419	2	A41182
38	290	3.1	1428	2	T08852
39	290	3.1	1838	1	CGHU1V
40	290	3.1	5262	2	T03454
C 41	288	3.1	1958	2	B40505
C 42	287.5	3.1	1464	1	CGHU1S
C 43	286.5	3.1	2715	2	T13049
C 44	284.5	3.0	1585	2	T31611
45	284	3.0	1151	2	T18535

ALIGNMENTS

RESULT 1

T31611

hypothetical protein Y5088A.g - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

A;Accession: T31611

R;Steward, C.

submitted to the EMBL Data Library, September 1999

A;Reference number: Z21047

A;Accession: T31611

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1585 <WIL>

A;Cross-references: EMBL:AL117200; NID:e1549770; PIDN:CAB55050.1; CESP:Y5088A.g

A;Experimental source: clone Y5088A

C;Genetics:

A;Gene: CESP:Y5088A.g

A;Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Alignment Scores:

Pred. No.: 6.67e-16 Length: 1585
Score: 382.50 Matches: 342
Percent Similarity: 27.70% Conservative: 84
Best Local Similarity: 22.24% Mismatches: 480
Query Match: 2 Indels: 632
DB: 2 Gaps: 57

US-09-931-704-3 (1-5087) x T31611 (1-1585)

Qy	6	GGAGTGGCGCTGGCGGATGGATTATTAAGCTTCGCGGAGCGCGCTCGCCCTCC 65
Db	541	SerGlyGlySerGlyGly-----GlyTyrSerSerAlaAlaProPro 556
Qy	66	ACTCCGCGAGCTCGCGGAGGAGGAGCCGCCGCCGCCAGCCGCCCA----- 120
Db	557	ProProProProProPro-----ProAlaProAlaProSerGl 572
Qy	121	-----TGGACCTCCGAGCAGGTTGAAACCCAACTAGCCCTGCTTTCATACATGACAA 176

QY	3191	TG	TATGCAGGTGACAGTGTCAAGCAGTGGC-----GGTGGTGCAGC	3153
Db	473	ly	ProlysGlygluIleGlySerAlaGlyProGlnGlyAlaProGlyProAlaGlyGlu	493
QY	3152	TC	CCAATCGAGAGGCGTA-----	3134
Db	493	lu	GlyLys--ArgGlyAlaAArgGlygluProGlyAlaAlaGlyProLeuGlyProG	512
QY	3133	-G	AAAAAGAGGAGGAGGAGAC-----GCACAAGCAGCTCAGGAAGA----	3092
Db	512	y	GluaArgGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGlyProLy	532
QY	3091	-----	CGTGCTGATVGGCTGGTGGGCAAAACCTTGGGAGAGTGCCCC	3051
Db	532	s	GlyAlaProGlyGluArgGlyValProGlyLeuGlyGlyProLysGlyGlyAsn----	550
QY	3050	GCT	TGTGGGGCAGGACAGAGAGAGAGCCCTGGA-----CGAGCTGGAGAGGTAGA	2997
Db	551	-----	GlyAspProGlyArgProGlyGluProGlyLeuProGlyAlaAlaArgGlyLeuTh	568
QY	2996	AG	TGCCCCCAGAGAGTGCAGAGACAAGAGAGAGGAGTTCGAGAAGCAGCAGCAGCTCA	2937
Db	568	x	GlyArgProGlyAspAlaGlyProGlnGlyLysValGlyProSerGlyAlaAlaGlyG	588
QY	2936	GCA	TGTGGCTGCAAAACAGACAGCAGAGAGAAATGAGCCITCAAAACACTGTGCGGCTTG	2877
Db	588	u	-----AspGlyArgProGlyProProGlyPro-----	597
QY	2876	GCA	AGGAGGTGCCTGACGACCTTGGAGAGAACCATTTGCTGGG--CCTGGAGACCAGAT	2820
Db	598	-G	lnGlyAlaArgGlyGlnPro-----GlyValMetGlyPheProGlyProLysG	614
QY	2819	TG	CAGGGGTTTAAGAAAGTGAAGTGGGTGGAGAGAAAGTGGAGCAGAGGGTGTAGACCACT	2760
Db	614	y	AlaAsnGly-----GluProGlyLysAlaGlyGluLysGlyLeuGlyGlyAlaProG	632
QY	2759	CC	TGGAGAAATTTGCGAGGAAAGCCAGGACAGAAAAAGGTCGGAACTAGAGAGGC	2700
Db	632	y	Leu-----ArgGlyLeuPro-----GlyLysAspGlyGluThrG	644
QY	2699	AG	TGGAGTGAAGCAAAAGTTTCTGCACAAGGAACAAGAGAGAGATGAGAGTGGCCCCATT	2640
Db	644	y	AlaGlnGlyProAsnGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlyProPr	664
QY	2639	GGA	ACAGAGCCCCATTTGCCGCTGACA--CACCCAGCCCCACAGCTAGAGGGCTGGG	2583
Db	664	o	GlyProSerGlyPheGlnGlyLeuProGlyProGlySerProGlyGluGly--G	683
QY	2582	GAG	ACGGGTACTCACTAACACTCCGAAGAGGGTTACGCTGATAGGACAGGGTTGC	2523
Db	683	y	LysProGly-----AspGlnGlyValProG	692
QY	2522	AGA	GGAAGAANAATGGAGGCAGTCTCATGATAGGACAGGAGAGGCTTATGGAAGACGC	2463
Db	692	y	GluaGlyAlaAlaGlyLeuValGlyProArg-GlyGluArgGlyPheProGlyGluA	712
QY	2462	AT	GTATATA-----GGGAAGAGGGCAGAGCCC-----	2437
Db	712	rg	GlySerSerGlyProGlnGlyLeuGlnGlySerArgGlyLeuProGlyThrProGlyT	732
QY	2437	-----	-----	2437
Db	732	hr	AspGlyProLysGlyAlaThrGlyProSerGlyProAsnGlyAlaGlnGlyProProG	752
QY	2436	-----	TTTCCAACCTCGGGGAGCGGCTGGGCCCCCAGGGAAGGATGAGGGA	2388
Db	752	ly	LeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyIleSerGlyProLysGlyAspA	772
QY	2387	GAT	TATATACTGGGAAGAA-----GGGGAAGCAAGGGGAGAGGCC	2346
Db	772	rg	GlyAspThrGlyGluLysGlyProGluGlyAlaProGlyLysAspGlySerGlyL	792

[illegible]

Db 1109 GlyAlaArgGlyLeuAsnGlyProGlnGlyProArgGlyAspGlyGlyGluAlaGlyGlu 1128
QY 1496 -----AGCGGTGAGTATAGG---TTTTCTGGATGAGGGGCGAGCCCTG 1452
Db 1129 AlaGlyGluArgGlyGlnLysGlyHisArgGlyPheThrGlyLeuGlnGlyLeu--ProG 1148
QY 1451 GGTCCCTGTGCGATTGAGAGCTGGCACTGCAGGAGGTGCCAGAGCACCGCTGCACAGC 1392
Db 1148 lyProPro-----GlyThrAlaGlyAspGlnGlyAlaThrGlySerAlaG 1163
QY 1391 ACCTAACATCCCCAC-----CAGTCCCTGTGGCAGGATGAGCAGAGCATGAGC 1338
Db 1163 lyProSerGlyProArgGlyProGlyProValGlyProSerGlyLysAspGlySerA 1183
QY 1337 GCAGC-----CGGGCCACACCGGGGAAGCA----- 1312
Db 1183 snGlylleProGlyProIleGlyProGlyProArgGlyArgSerGlyGluThrGlyP 1203
QY 1311 --GCTGGCAGCCCC-----TCAGGTCTCTG-----TCCCCACCTTGGGCTGGGT 1270
Db 1203 roAlaGlyProGlyGlnSerGlyProGlyProGlyProPro-----GlyP 1221
QY 1269 GGGGGTGTGGGTGGGACACAGATGGAGGACAGGGATGGAGGGGGAACACATCC 1210
Db 1221 roGlylleAspMetSerAlaPheAlaGly-----L 1231
QY 1209 TGGCAAGACGGGAGAAAGAGAGGCATGTGTATGTTTGGAGTATGAGCATGGAGCTGGCAGCA 1150
Db 1231 euSerGlnProGluLysGlyProAspProMetArgTyrMetArgAlaAspGlnAlaSer 1251
QY 1149 TGAGTGTCTATAGTATGATCTTGGAGATGATCTGTGTGTGTCAGCTGAGAGAGTCTTTGG 1090
Db 1251 erSerValleuGlnThrAlaAspValGluAlaAsnLeuLysThrLeuAsnAsnGlnIleG 1271
QY 1089 CATGTGTG-----CCACTGTGTGTGTGCCAACATG 1060
Db 1271 luSerllelleArgSerProAspGlyThrLysLysAsnProAlaArgThrCysArgAspL 1291
QY 1059 TATATGTGTG-----ACATGAGGAGGACGGGCGGCGCTAGGAGCTGCGAGTCCGACCT 1006
Db 1291 euLysLeuCysHisProGluTrpLysSerGlyGluTyrTrpIleAspProAsnGluGlyC 1311
QY 1005 GC-----CTGTGTCCAGTTTACCTAGCTTCCTGCTCCGAGCTTT 967
Db 1311 yThrValAspAlaIleLysValPheCysAsnMetGluThrGlyGluSerCysValTyrP 1331
QY 966 CCACTCCCACT 956
Db 1331 roSerProSer 1334

RESULT 3

S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
R:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Alignment Scores:

Pred. No.: 4,97e-13 Length: 1188
Score: 336.00 Matches: 278
Percent Similarity: 30.50% Conservative: 88
Best Local Similarity: 23.17% Mismatches: 321
Query Match: 3.56% Indels: 516

Db: 2 Gaps: 50
US-09-931-704-3 (1-5087) x S49915 (1-1188)
QY 15 CTTGCGGATGGGATATTAAAGCTTCGGCGGAGCGCGCTCGCCCTCCCACTCCGCCCA 74
Db 412 ProGlyGly-----GlyProProSerSerPro 420
QY 75 GCCTCGGAGAGGAGCGCACCCGCGC---CGGCCAGCCCC---AGCCCCATGGACCTC 128
Db 421 ValProGlyLysProAlaAlaSerAlaProMetProSerProHisThrProAspVal 440
QY 129 CGAGCAGGTGAAACCCAACTAGCCCTGTCTTCTAATACATGACAAGCAGCGCCCAT 188
Db 441 SerProGluProLeuProGluProSerProVal-----ProAlaProAla 455
QY 189 CTGATACCTAAACCCACCAAGTCACAGC-----CCTCAACTCA 227
Db 456 ProMetProMetProThr-ProHisSerProProAlaAspAspTyrValProProThrPr 475
QY 228 CCTCTGCTGCCAGAGCTCACCA----- 252
Db 475 oProValProGlyLysSerProProAlaThrSerProSerProGlnValGlnProProAl 495
QY 253 -----CATCCTTTGGACT 266
Db 495 aAlaSerThrProProProSerLeuValLysLeuSerProProGlnAlaProValGlySe 515
QY 267 CAACCTCAACGGCACTAAATCAACAATCCAAAGTCTAAACTAATCTGAAACTTTTAA 326
Db 515 rProProProPro-VallysThrThrSerPro----- 525
QY 327 AGTAACCCAGTCTTAAACCTAGCCCAATGCCAATTATATCTACCTAGCCAAAC 386
Db 526 -----ProAlaProIleGlySerProSerProProProValSerValSerPro- 543
QY 387 CTAACCTCCTTTGCCAGTCCAAAGTGTCCACTGAATCTCCTTGGTCTCCTCACTGAAA 446
Db 544 -----ProProProValLysSer-ProProProProAlaProValGlySerProPro 560
QY 447 ATCCAGAAAGCATATTTCCCACTGCCACATCCCTCTTACAGCAGCCACCTTGGC 506
Db 561 ProProGluLys-----SerProProProAlaProValAlaSerProProProPro 578
QY 507 CTTGGACT-----CCTGGTATCTGGGATGTCCAAACTCTGCAGTGCATCAGCC 557
Db 579 VallysSerProProProProThrLeuValAlaSerProProProProValLysSerPro 598
QY 558 AACAGCCGAGCTCGTCAATGCACTCTCTCCCTTCTGTCCCACTTGCAGGCTGA 617
Db 599 ProProPro-----AlaProValAlaSerPro-ProProProValLys----- 612
QY 618 TGGAAAGGCCTCATTTGAAGTCCAACTTTTCCCACTAACACACAAAGAGGGGTGAACCT 677
Db 613 -----SerProPro-ProPro--ThrProValAlas 622
QY 678 CCACACTGCCACCGTTCCTGAGAGTGAGCACTAAATCTCTTCAATCTAAACCCACCT 737
Db 622 erProProProProAlaProValAla-----SerSerProProPro 636
QY 738 ACATTCCCACTCAGGAATCATCTAGTAATATACCAAACTAAGCCCAATAGGC 797
Db 636 etLysSerPro-----ProProProThrProValSer 647
QY 798 AGCCGAGCCCTAGTGGTCTTAACCTATACCTTGTCTTCTATGGTGAGTCTGTCTTGGC 857
Db 647 erProProProProGluLys-----SerProp 656
QY 858 GGCGGCTCTCTCTCTGCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAC 917
Db 656 roProPro-ProProAlaLysSerThrPro----- 665
QY 918 ATGTGTGTCTCCCACTCTGACTCTCCCTCTCAAGCTGCAGTGGAGTGGAAAGACTGGCAG 977

666	Db	: : : -----ProProGluGluTyrPro-----	671
978	Qy	GAAGCTAGGTACAACTGGAACACAGCGCAGGTGCACTGCAGTCCCTAGSCCTGGCCCCG	1037
671	Db	-----	671
1038	Qy	TCCCTCCATGTACACACATATACATGTTGGCACACACAGTGGCACACATGCCAAAGAC	1097
671	Db	-----	671
1098	Qy	TCTCTCAGCTGACACACAGATCCATTCTCAAGTATCTACTGATAGACACTCATGCGTGCC	1157
672	Db	: : : -----ThrProThrSer-----	676
1158	Qy	AAGTCCTCATCTCAACATACACATGCCTCTCTTCTCTCCGCTTTGCCAGGAGTGT	1217
677	Db	: : : -----VallySerSer-----	680
1218	Qy	TCCCTCTCCATCCCTCTGCTCTCCCATCTGGTCCCACTCTCAC-CCCCACCCAGC	1276
681	Db	: : : -----ProThrLeuLeuProSerProPro-----	697
1277	Qy	CAAGTGGGGACACAGACTGAGGGGCTGCCAGCTGTTCCCGTGTGGGCC-----	1330
698	Db	: : : ProGlnGluYsProThrProProSerThrProSerYsProProSerSerProGluYs	717
1331	Qy	-----GGCGCGCTCATGCTTCTCGTCCATCTCGCCACAGGGACTCGTGGGGATG	1384
718	Db	: : : ProSerProProYsGluProValSerSerProProGlnThrProYsSer-----	734
1385	Qy	TTAGCGTCTGTGCACGCTGCTCTGGCACCTCTCCCTGCAGTGCAGCTCTCAATCGCAC	1444
735	Db	: : : -----SerProProAlaPro-ValSerSerProPr-----	745
1445	Qy	GGGGACCAAGGCGTGCGCCCTTCATCCAGAAAACCTTANGACTTCACCGCTACCTGGAG	1504
745	Db	: : : oProThrProValSerSerProProAlaLeuAlaProValSerSerProSer-----	763
1505	Qy	CACCACTCCGCGACTTGCGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGGGAGTT	1564
763	Db	-----	763
1565	Qy	GGGGAGGAGTGAGAGTTGGGGAAGACAGTCTCAACCGTGGAGGGTTCTGTAATGAT	1624
764	Db	: : : -----VallySerSerPro-----	768
1625	Qy	GGGGTGAGGAGGGGCTTTTGGCTCCACAGTCCCTGTCTGGTCTATCTCTGCGCT	1684
769	Db	: : : -----ProProAlaProLeuSer-----SerPro-----	776
1685	Qy	TCCCTCTTAGTGGCCCCCACTTCCCATCTCCCTGGCCCCCAGGACTAGGATGTGGCA	1744
777	Db	: : : -----ProProAlaPro-----	781
1745	Qy	GGCTCGCACCCGCTTGCCCATTTGCCCACTGCTGCCAGCCAGCCCGCCCTCCC	1804
781	Db	: : : nVallySerSerProProProValGln-----ValSerSerPro-ProProAlaProL-----	799
1805	Qy	CTTGGGGCGGGGAAGTCTCTCTGTTTTCACCGTGTGTGGTGTCTCTTTCGCGGGCG	1864
799	Db	: : : ysSerSerProProLeuAlaProValSerSerPro-ProGlnVal-----	813
1865	Qy	GGGTGGTGGGACAGAGGGGCCCCCATCTCCATGCTGGTTCACAGTCTGCTCTGCC	1924
814	Db	: : : -----GluYsThrSerProProProAlaProLeu-SerSerProProLeuAl-----	829
1925	Qy	CCAGACCTGGGGCCCTGTGCTCTGGACCCAGGGGCGCTCCCTTCGCTCGCTCTCCCA	1984
829	Db	: : : aProYsSerSerPro-----ProHisValValValSerSer-----	841
1985	Qy	TCCTAGCTGGGCTCTCTAGGGGGGTGATGGGGGAAGGGGACTGTAGGGGAACCCAGGCGAT	2044
		: : : -----	

Db	841	rPro	-----ProProVa	845
Qy	2045	AGTGGCAGGGGTTTAGGGTGTGGATGGAGGTATATGCTGAAGGATTGGGGGTGGTCCA	2104	
Db	845	lVal	-----	846
Qy	2105	GAGGTGTTCAGAGACCCAGGAGAGAGAGAGAGAGGAGGTGGAGGAGCCGAGGCACCATGG	2164	
Db	847	-----LysSerSerPro	-----	850
Qy	2165	GGAAACGGCCCTCTCCCGTGTTCCTTCCATCCACAGACCCCTACTCTGGAGCCAG	2224	
Db	851	-ProProAlaProValSerSerProProLeuThr	---ProLysProAlaSerProProAl	869
Qy	2225	GGAAAGAAAGGAAGGTGGCGGGAGCTGGCTCCAGCCGCCAGATACACCGAGGA	2284	
Db	869	aHisValSerSerProProGluValValLysProSerThrProProAlaProThrVa	889	
Qy	2285	AATTAGTTGTCCTGTGCTGTGTGAGCGTGTGAACCTCCCGCTGGGCCCTT-GCCTATCC	2343	
Db	889	lileSer	-----ProProSerGluProLysSerSerPr	900
Qy	2344	CAGGCTCTCCCTTGCTTCTCCCTTCTTCCAGTTATATACATCTCCCTCATCCCTTCC	2403	
Db	900	oProProThrProValSerLeuPro	---ProProIleValLysSerSerProProPr	918
Qy	2404	CTGGGCCCGAGCGCTCCCGCAGGGTGTGAAAGGGCTGTGCCCTCTTCCCTATACCATG	2463	
Db	918	oAlaMetValSerSerProProMetThrProLysSerSer-Pro	-----	932
Qy	2464	CTGCTCTTCATAGCCTTCTCTGTCTCTACTCATGAGACTGCCTCCATTTCTTCCCTCTG	2523	
Db	933	-----ProProValValValSerSerProProThrValLysSerS	947	
Qy	2524	CAACCTGCTCTCATCAGCTGAACCTTCTTTCGGAGTGTGTAGTAGTACCCGCTCTCTCC	2583	
Db	947	er	-----ProProProAlaProValSerSerProProAlaThrP	960
Qy	2584	CCAGCCCTCAGCTGTGGCGCTGGGTGTGTGACGGCAATGGGCTCTGGTTCCAATG	2643	
Db	960	roLys	---SerSerProProAlaProVal-----AsnL	971
Qy	2644	GGCACTCTCATCTCTCTCTTCTGTGTGTCGAGAAACCTTTGCTTCATCTCCACTGCC	2703	
Db	971	euProProProGluValLysSerSer	-----ProPro-----ProThrProV	985
Qy	2704	TCCTAGTTCCCGACCTTTTCTCTCTGGCTTTCCTGCCAATTTCTCCAGGAGTG	2763	
Db	985	alSerSerProProPro	-----AlaProLysSerS	995
Qy	2764	GTCTACACCTCTGCCT	-----CCACTTCTCTC	2790
Db	995	erProProProAlaProMetSerSerProProProGluValLysSerProProProp	1015	
Qy	2791	TCCACCCACTCACTCTTAACCCCTGCAATGTGGCTTCAGGCCCCAGCAATGGTCTC	2850	
Db	1015	roAlaProValSerSerProProProProValLysSerProProPro-----AlaP	1033	
Qy	2851	TCCMAGTGTGTAGGCACTCTCTGCAAGCCCGACGTGTTTGAGGCTCATCTCTCT	2910	
Db	1033	roValSerSerProProProValLysSer	-----	1043
Qy	2911	TGCTGTCTGTTTGAGCCACACTGCTGAGCGCTCTGCCTTCTCGAACTCCTCTCTCT	2970	
Db	1044	-----ProProProProA	1048	
Qy	2971	GGTCTGTGCACTCTCTGGGCCACCTTCTACCTCTCCAGCTCTTCCAGGCTCTCTTCT	3030	
Db	1048	laProValSerSerProProProProValLysSerProProProProAlaProIleSerS	1068	
Qy	3031	CTGTGCTCTGCCCCACAGCGGGCACTCTCCCAAGGTTTGCCACCCAGCCMATCAGCAC	3090	
Db	1068	er	-----ProProProProValLysSerP	1076

[illegible]

QY 3800 TCTTGAGCCGGTTGAAGTCTCTGGCCGAGCGCCACAGCCAGGCTCTGCAGCTCTCTTCAGCA 3741
Db 364 -----SerAlaGlyProSerGly 369
QY 3740 GCACAGAGTCGTCCATCTCTGGAGGAAGTCACTGTGGGAGGCGCCAGGAGTCCAAAGTGG 3681
Db 370 AlaGlnGlyThrProGlyPheProGlyAlaAargGlyProProGlyProAlaGly----- 387
QY 3680 GTTCAGTCCACGAGCGGCTGGGCGAGTGGGTAGCCAGAGCTGCCATGACGCCCGCAA 3621
Db 388 -----SerProGlyProAlaGlySerLysGlyAsp 398
QY 3620 TGCTGCCAGCAGCCCTCGAGGCTGGTGCAGAAAGTGGCCAGGCTGGCGGCGAGCTCAG 3561
Db 399 GlyAsnProGlyGlnProGlyAlaGlnGlyLysGlyPro---LeuGlyProArgGly 417
QY 3560 CAGTGGCAGCTGACGGTTGAGGCCACGCAAGTAACAGAGTGGCTGTAGGCTCGT 3501
Db 418 GluThrGlyProAlaGly-----ProProGlyAlaGlnGlyGlySerGlyGluArg 434
QY 3500 AGTTCTGGGTCAGCGGAGTTGTGATTGAGGCTTCGCCACACCTCCAGAGTCAACAGTGG 3441
Db 435 GlySerArgGlyAlaLeu-----GlyProAlaGlyProProGlyGlyValGly 450
QY 3440 CCTGGGCAGAGTCTCTGCCCCAGGCGGGAGGGTTGAAGTCTGGCTCGTTGAAAGGGG 3381
Db 451 GluArgGlyProMetGlyProProGlyMet-----SerGly 462
QY 3380 GGCACAGTAGTTCAGCTGTGA--AGAGAGGGGTATGG 3342
Db 463 AlaProGlyAlaProGlyAlaLysGlyAspArgGlyLeuProGlyGluArgGlySerAla 482
QY 3341 GGAAGG-----AGGAGGCGAGGCGCTGGCTCACCACCAA 3306
Db 483 GlySerLysGlySerAlaGlyGlySerGlyArgProGlyGluPro---GlyMetProGlyG 502
QY 3305 GGAGATACCTGCTCCCAAGTTCTATTTTGTGTCCTCGACACTGGCCCTGCTCCA 3246
Db 502 LnaArgGlyLeu-----ThrGlyProProGlyL 511
QY 3245 TGCTAGGCTCTTGTTCATGGCTCCCTGAGTTTTCATAGGATATCATTTGTATG 3186
Db 511 ysGlnGly-----ArgAspGlyLysPro-----GlyProAlaGlyAlaP 524
QY 3185 CAGGTGACAGTTCAGAGCAGTGGCGGTGGTGCAGCTCTCCATGG-AGAGGCGTAGAAAAAG 3127
Db 524 roGlyGluProGlyAsnSerGlyProAlaGlyAlaSerGlyGlnArgGlyLeuProGlyL 544
QY 3126 GAGGAGGAGGAGGACAAAGAGCTCAGGAAGACGTGTGATGCTGGGTGGGCAAA 3067
Db 544 euValGlyLeuProGlyProGlnGlyGlnArgGlyGluArgGlyGluAspGlyGlyGlnG 564
QY 3066 CCTTGGGAGAGTCCCG---CTGGGGGCGAGCAGACAGAGGAGAGAGGAGCTGGAGAG 3010
Db 564 lysSerProGlyAlaProGlyLeuThrGlyGluProGlyLysArgGlyGlyProGlyVala 584
QY 3009 CTGGAGAG-----GTAGAAAGTGGCCAGGAGAGTGCAGACCAAGAGGAGAGTTC 2956
Db 584 laglyProProGlyProGlnGlySerAlaGlyGluArgGlyAsnGlnGlyProGlnGlyG 604
QY 2955 GAGAAGGCAGC-----ACGGTTCAGCATGTGTGGGTGCAAAA 2920
Db 604 lnaAlaGlySerMetGlyProProGlyProGlyAlaSerGlyAspAlaGlyAlaGlnG 624
QY 2919 CAGACAGCAAGGAGATGAGCCCTTCAAAACACTGTCCGGCTTGCAGAGAGGTCCCTGAC 2860
Db 624 lysAsn----- 626
QY 2859 GACCTTGGAGAGAACCATTTCTGGGGCTTGAAGCCAGATTGACAGGGGTTAAGAGTGA 2800
Db 627 -----GlyProProGlyGlu-----S 632

QY 2799 GTGGGTGGAGAGGA---AGTGGAGGCAGAGGGTGTAGACCCTCTTGGAGAAATTTGGC 2743
Db 632 erGlyProGluGlyProProGlyAlaAargGlyGluArg----- 644
QY 2742 AGGAAAGCCAGGAGAGAAA-----AAGGTGGGAACCTAGAGAGGCGAG 2698
Db 645 --GlyAlaProGlyGluAargGlyProGlnGlyLeuThrGlyAlaGlnGlyArgArgGlyI 664
QY 2697 TGGAGTGAAGCAAAAGTTTCTGCACAAGGAACAAGAGAGATGAGATGGCCCATGG 2638
Db 664 leglyGlyArgAlaGlyAsnTyrGlyAlaThrGlyProProGlyGlnLysGlyGluMetG 684
QY 2637 AACCAAGCCCCATTTGCGCTGCACACACCCAGGCCACAGCTGAGGGCTCGGAGAG 2578
Db 684 lyProProGlyAsnValGlyLeu-----GlnGlyProProGlyGlu---LeuGlyProS 701
QY 2577 ACGGTACTCTAATACACTCCGAAAGAGGGTTCAGT-----GATAGGA 2533
Db 701 erGlyProProGlyAlaAargGlyProProGlyProSerGlySerProGlyProAspGlyP 721
QY 2532 GCAGGTTGCAGAAAGAAATGAGGCGAGCTCATGATAGTAGGACAGGAGGAAGGCTAT 2473
Db 721 roAlaGlyAlaGluGlyAspArgGly-ProValGlyProMetGlyProSerGlyProSer 740
QY 2472 GGAAGACAGCATGTATAGGGAAGAGGCGAGAGCCCTTTCCAAACCTCGGGGAGCG--- 2416
Db 741 GlyMetProGlyGluAargGlyAspAsnGlyGluProGlyPro-----GlyGlyAlaPro 758
QY 2415 -----GCTGGGCGCCAGGGAAGGATGAGGAGATGTATA 2380
Db 759 GlyGlnArgGlyAspProGlyAspLeuGlyProGlnGlySerProGlySerProGlyPhe 778
QY 2379 ACTGGAAAGAGGAGAGCAAGGAGGAGGC-----CTGGATAGGCAAGG 2332
Db 779 AlaGlyProProGlyArgSerGlyAsnProGlyProGlnGlyGluLeuGlyProThrGly 798
QY 2331 GCCAGGGGAGGTTTCACACGCTGACAAGACAGACAGACAACTAATTTCTCGGTGTATC 2272
Db 799 AlaArgGlyGluThrGly----- 804
QY 2271 CTGGGCTGGAGCCAGCT-----CCCCCGCCACCTTCTTCCCTTTCTTT 2227
Db 805 ---GlyProGlyProSerGlyProThrGlyAspProGlyProGlnGlyProLeuGlyAla 823
QY 2226 CCCTGGCTCCAGAGTAGGCTCTGGATGGAAGAGGACACGGAAGAGGCGCGCTT 2167
Db 824 ProGlyGlnGlnGly-----GluArgGlyGluThrGly----- 834
QY 2166 CCCATGTGCTCGCTCGCTCCAAACCTCTTCTTCTCTCTCTGCTGGCTCTCTGAACACC 2107
Db 835 ProGlnGlyGlnGlyGlyPro-----ProGlyProIleGlySerLeuGlyAla 850
QY 2106 TCTGGACCA-----CCCCAAATCCTTACACATAACTCCATCCACACCCTAAACCCC 2053
Db 851 ProGlyAlaGlnGlyProProGlyPro-ThrGly-----ProSer----- 863
QY 2052 CTGCCACTACTGCTGGGTTCCTCAGTCCCTTCCCTCCCATGACCCCCCTAGAGGCC 1993
Db 864 -----GlyAsnAlaGlySerProGlyGlnProGlyAlaAargGlyGluPr 878
QY 1992 AGCTAGTAGTGGAGAGGCGAGC-----GGAAGGAGGCCCT---GGTC 1951
Db 878 oGlyGlnSerGlySerProGlyGlnProGlyLeuAlaGlyA-gThrGlyProSerGlyG 898
QY 1950 CAGAGCAGAGGCGCCAGGTTCTGGGGCAGAGGCGAGCTGGAACGACGATGGGAGGT 1891
Db 898 uArg-----GlyAspLysGlyAsnAspGlyGlnSerGlyProPr 911
QY 1890 GGGGCCCTCTGTCTCCCAACCCCAACCCCGCGCAAGAGACACACACAGCTGTAA 1831
Db 911 oGlyPro-----ProGlyProAlaGlyProAlaGlyGlnSerGlyIleLeuGlyLeuAl 929
QY 1830 CAGAGGAGACTTCCCC-----GGCCCCAGGGGAGGCGGGC-----GGCTGGCTGGCAG 1780

QY 576 AATGCACTCTCTCCCTCTCTGTCCTCCACCCCTTGACGCTGATGGAAGGCTCATTTGAA 635
Db 1735 MetAlaThrMetSerThrIleHisProSer----- 1744
QY 636 GTCCAACTTTTCCCACTTAACACCAAGAGGGGTGACCTCCACACTGCGACCGTTCC 695
Db 1745 -----SerThrProGluThrThrHisThrSer 1753
QY 696 CTGAGAGTGAGCACTAAA-----TCTCCTTCA 722
Db 1754 ThrValLeuThrThrLysAlaThrThrArgAlaThrSerSerMetSerThrProSer 1773
QY 723 ATCTAACCCCACTTACCTTCCACACTCAGCAATCATCTCAGATATATACCAAAAC 782
Db 1774 SerThrProGlyThrThrTrpIleLeuThrGluLeuThrAlaAlaThrThrAla 1793
QY 783 TAAGCCCAATAGGACGCGCGACC-----CTAGTGGTCTAACCC 821
Db 1794 AlaLeuProHisGlyThrProSerSerThrProGlyThrThrTrpIleLeuThrGluPro 1813
QY 822 TATACCTTGCTTCTATGGGTGAGTCTGTTCTTGGCGCGCCCTCTCTCTCTGCTTCTCC 881
Db 1814 SerThrAlaThrVal-----ThrValProThrGlySerThrAlaThrAlaSerSer 1831
QY 882 CTTAGAGCTAGCTGCTCAGCTGCGACGCTGACAGTGTCTGCTCCACCCCT----- 936
Db 1832 ThrArgAla-ThrAlaGlyThrLeuLysValLeuThrSerThrAlaThrProThrVa 1851
QY 937 -----CTGACTCCCTCAAGCTGACGTGAGCTGGAAGA-----CTGG 974
Db 1851 lIleSerSerArgAlaThr-ProSerSerSerProGlyThrAlaThrAlaLeuProAlaL 1871
QY 975 CAGGAAGCTAGGTACAACTGGAACACAGGCGAGGTGACCTGACCTGCTCCCTAG----- 1026
Db 1871 euArgSerThrAlaThrThrProThrAlaThrSerValThrAlaIleProSerSerL 1891
QY 1027 -----GCTGGCCCCCTCCCTCCATGACACA-----CATATACATGTTGGACAC 1072
Db 1891 euGlyThrAlaThrThrArgLeuSerGlnThrThrProThrAlaThrMetSerThr 1911
QY 1073 ACACAGTGGCACACATGCCAAAGACTCTCTCAGCTGACACACACATCCATCTCAAGTAT 1132
Db 1911 laThrProSerSerThrProGluThrValHisThrSerThrValLeuThrThrAla 1931
QY 1133 CTACTGATGACACTCATGCGTG-----CAAAGTCTCATCTCT----- 1170
Db 1931 hrThrThrArgThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThr 1951
QY 1171 -----CAAACATACATGCTCTCTTCTCTCCGCTTCCAGGAGTGTTCCTCC 1222
Db 1951 hrLysValProThrThrThrThrThrGlyPheThr-----AlaThrPro--SerSerSerPr 1969
QY 1223 TCTCCATCCCTCTGCTCCCTCTGCTGCTGCCACCTCACCCTCACCCTCCAGCCCAAGG 1282
Db 1969 oGlyThrAlaLeuThrProProValTrpIleSerThrThrThrThrProThrThr 1987
QY 1283 TGGGGACAGACACTGAGGGGCTGCCAGCTCTCCCGTGTGGGCGCGCGCTCA 1342
Db 1988 -----ArgGlySerThrValThrProSerSerIleProGlyThrThrHi 2002
QY 1343 TGCT----- 1354
Db 2002 sThrAlaThrValLeuThrThrThrThrThrThrValAlaThrGlySerMetAlaThrPr 2022
QY 1355 ATCTCTCCCAAGGACTCGTGGGGATGTAGCGTGTGACGGTGTCTGCGAC 1414
Db 2022 oSerSerSerThrGlnThrSerGly-----ThrProPr 2033
QY 1415 CTCCCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGGGCTGCCCTCCATCCAG 1474
Db 2033 oSerLeuThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProSerSe 2053
QY 1475 AAAACCTATGACCTCACCC-----GCTACCTGGGAGCACCACTCCGAGC 1519

Db 2053 rThrProGlyThrThrProIleProProValLeuThrThrThrThrThrProAlaA 2073
QY 1520 TTGGCTGGGACCTATGTAGTATCCAGCCTAGGAATCTGGGAGTTGGGAGGAGTAGGA 1579
Db 2073 aThrSerSerThrValThrProSerSerAla-----LeuGlyThrThrHis-- 2088
QY 1580 GTTGGGGAAGACAGTCTCTAA-----CCGTGGAGGGTCTGTAAATGATGGGTGAGGAG 1635
Db 2089 -----ThrProProValProAlaThrThrAlaThrThrHisGly-- 2101
QY 1636 GGGCTCTTTGGCTCCCAACAGT-----CCCTCTGTCTGTCTATCTCTGCTCC 1683
Db 2102 -ArgSerLeuProProSerSerProHisThrValProThrAlaThrThrSerAlaThrSe 2121
QY 1684 TTCCCTCTTAGT-----GGCCCCCCTCTCCCATCTCCCTGCCCCCAGG 1728
Db 2121 rGlyIleLeuGlyThrThrHisIleThrGluProSerThrGlyThrSerHisThrProAl 2141
QY 1729 A-----CTAGCATGTGGCAGGCTCCACCCGCTTGGCC-----CATTTGCCCCAC 1776
Db 2141 aAlaThrThrGlyThrThrGlnProSerThrProAlaLeuSerSerProHis--ProSe 2160
QY 1777 TGGCTGCCAGCCAGCCGCTCCCTCCCTGGG-----GGCCGGGG 1818
Db 2160 rSerArgThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrArgL 2180
QY 1819 AAGTCTCTCTGTTTACACCGTGTGTGTCTCTTGGCGCGCGGG-----TTGGG 1872
Db 2180 yThrSerArgThrThrAlaThrAlaThrProSerLysThrArgThrSerThrLeuLeuPr 2200
QY 1873 TGGGACAGAGGGGCCCTCCATCGCTGCTTCCAGCTCGCTCTGCCCCCGACAC 1932
Db 2200 oSerSerProThrSerAlaProIleThrThrValValThrThrGlyCysGluProGln-- 2219
QY 1933 TGGGGCCTCTGCTCTGACCCAGGGGCTCTCTTCCCTCTGCTCTCTCCATCTAGCT 1992
Db 2220 -----CysAlaThrSerGluTrpLeuAspTrpSerLysProMetPro----- 2233
QY 1993 GGGCTCTCTAGGGGTCATGGGGAGGGAGCTGTAGGGAACCCAGGAGTAGTGGCAG 2052
Db 2234 -GlyProSerGlyGlyAspPhe-----AspThrTyrSerAsnIleArgAlaAlaGly-G 2251
QY 2053 GG-----GCTTAGGCTGT-----GGATGGA 2073
Db 2251 lAlaValCysGluGlnProLeuGlyLeuGluCysArgAlaGlnAlaGlnProGlyValP 2271
QY 2074 GGTTA-----TGCTGTAAAGATTTGGGGTGTGCTCCAGAGGT 2109
Db 2271 roLeuArgGluLeuGlyGlnValValGluCysSerLeuAspPheGlyLeuValCysArgA 2291
QY 2110 GTTCAGAGAGCCAGGAGAGGAAGG-----AGGTTGGAGGAG 2150
Db 2291 snArgGluGlnValGlyLysPheLysMetCysPheAsnTyrGluIleArgValPheCysC 2311
QY 2151 CCGAGGACCACTGGGGAACCGGCCCTCTTCCCGTGTCTCTTCCATCCATCCAG----- 2206
Db 2311 ysAsn-----TyrGlyHisCysProSerThrProAlaThrSerThrAlaThrProS 2329
QY 2207 -----ACCTACTCTGGAGCCAGGGAAGGAGG 2237
Db 2329 erSerThrProGlyThrThrTrpIleLeuThrLysLeuThrThrThrAlaThrThrG 2349
QY 2238 AAGAAGGTGGGGAGCTGGCTCCAGC-----CCAGATACACCGAG 2282
Db 2349 luSerThrGlySerThrAla-ThrProSerSerThrGlnGlyProProAlaGlyThrPro 2368
QY 2283 GAAATTAGTTTG-----TCTCTGCTCTGTCTAGCGTGTGAACCTCCCCCT 2327
Db 2369 HisValSerThrAlaThrThrProThrValThrSerSerLysAlaThrProPheSer 2388
QY 2328 GGGGCC-----TTGCCTATCCAGCCCTCTCCCTTGTCTTCTCTTCTT 2372

Db 2389 SerProGlyThrAlaThrAlaLeuPro-AlaLeuArgSerThrAlaThrThrProThrAl 2408
Qy 2373 TCCAGTATACATCTCCCTCATCCCTTTCCTGGGCCCCAGCGCTCCCGAGGGTTG 2432
Db 2408 aThrSerPheThrAlaLeuProSerSerSerLeuGlyThr-----Tr 2423
Qy 2433 GAAAGGGCTCTCCCTCTCCCTATACCATCTGCTCTCCATAGCCTTCTCTCTCTCTA 2492
Db 2423 pThrArgLeuSerGlnThrThrProMetAlaThrMetSerThrAlaThrProSerSe 2443
Qy 2493 CTCATGAGACTGCTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2552
Db 2443 rThrProGluThrValHis----- 2449
Qy 2553 TTTCGGAGTGTAGTACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2612
Db 2450 -----ThrSerThrValLeuThrThrAlaThrThrThrThrGlyAlaThrG1 2465
Qy 2613 GTCAGCGGCAATGGGCTCTGGTTTCCCAATGGCCACCTCTCATCTCTCTCTCTCTCTCT 2672
Db 2465 ySerValAlaThrProSerSerThrPro--GlyThrAlaHisThrThrLysValProTh 2484
Qy 2673 TGCAGAAACCTTGCTCTCACTCCACTGCTCTCT-----CTAGTTCC 2714
Db 2484 rThrThrThrThrGlyPheThrValThrProSerSerSerProGlyThrAlaArgThr-P 2504
Qy 2715 CGACCTTTTCT 2774
Db 2504 roProValTriPileSerThrThrThrThrThrThrThrThrThrThrThrThrThr 2524
Qy 2775 CTGCC-----TCCACTTCTCTCTCCACCCACTCTCTCTTAACC-----CCCTGCA 2819
Db 2524 erSerIleProGlyThrThrHisThrProThrValLeuThrThrThrThrThrThrThr 2544
Qy 2820 ATCTGGCTTCAGGCCCGCCAGCAATGGTTCTCTCCAGGTGTCGAGCAGCTCTCTGCCAA 2879
Db 2544 laThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThr-----P 2561
Qy 2880 GCCGACAGTGTTTGAAGGCTCATCTCTCTGCTGCTGTTTTCAGCAGCAGCTGCTGA 2939
Db 2561 roProSerLeuLeuThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProS 2581
Qy 2940 GCGCTGCTGCTCTCGAACT 2978
Db 2581 erSerThr-ProGlyThrThrProIleProProGluLeuThrThrAlaThrThrPro 2600
Qy 2979 -----CACTCTCTCTGG 2990
Db 2601 AlaAlaThrSerSerThrValThrProSerSerAlaLeuGlyThrThrHisThrProPro 2620
Qy 2991 CCACCTCTTACCTCTCCAGCTCTCTCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3047
Db 2621 ValProAsnThrThrAlaThrThrHisGlyArgSerLeuSerProSerSerProHisThr 2640
Qy 3047 ----- 3047
Db 2641 ValArgThrAlaTriPThrSerAlaThrSerGlyThrLeuGlyThrThrHisIleThrGlu 2660
Qy 3048 -----AGCGGGCACTCTCCCAAGTT-----TGC 3071
Db 2661 ProSerThrGlyThrSerHisThrProAlaAlaThrThrGlyThrThrThrThrThrThr 2680
Qy 3072 CCACCCAGCAATCAGCAGCT 3128
Db 2681 ProAlaLeuSerSerProHisProSerSerArgThrThrGluSerProSerProGly 2700
Qy 3129 TTTTCTACGCTCTCAITGGAGAGCTCACCCGCACTGCTTCAACTGT----- 3179
Db 2701 ThrThrThrProGlyHisThrThrAlaThrSerArgThrThrAlaThrAlaThrProSer 2720
Qy 3180 CACCTGCATACAAATGATATCTTATTGAAAA----- 3212
Db 2721 LysThrArgThrSerThrLeuLeuProSerGlnProThrSerAlaProIleThrVal 2740

Qy 3213 ---ACTCAGGAGGAGCCATGAACAAAGAGCCTAGCATGGAGCAGGGC----- 3257
Db 2741 ValThrThrGlyCysGluProGln-----CysAlaTriPThrLeuAspThrSer 2758
Qy 3257 ----- 3257
Db 2759 TyrProMetProGlyProSerGlyAspPheAspThrThrThrSerAsnIleArgAlaAla 2778
Qy 3258 -----CAGTCTCAGGAGCAGCAAAAA----- 3278
Db 2779 GlyGlyAlaValCysGluGlnProLeuGlyLeuGluCysArgAlaThrAlaGlnProGly 2798
Qy 3279 -----ATAGAACTTTGGGAGCAGGTATC-----TCCTTGTGTGTG 3314
Db 2799 ValProLeuGlyGluLeuGlyGlnValValGluCysSerLeuAspPheGlyLeuValCys 2818
Qy 3315 AGCCAGCGG----- 3323
Db 2819 ArgAsnArgGluGlnValValGlyPheLysMetCysPheAsnTyrGluIleArgValPhe 2838
Qy 3324 -----CTCTGCCCTCCT-----CCTTCC 3341
Db 2839 CysCysAsnTyrGlyHisCysProSerThrProAlaThrSerSerThrAlaMetProSer 2858
Qy 3342 CCATCACCT 3377
Db 2859 SerThrProGly-----ThrThrTriPLeuThrGluLeuThrThrThr 2873
Qy 3378 -----GCCCTCTTCAACGAGCAGCACTTCA 3404
Db 2874 AlaThrThrThrAlaSerThrGlySerThrAlaThrProSerSerThrProGlyThrAla 2893
Qy 3405 ACCCTCCCGCTGGGGCAGAGCTCTGCCAGGGCAGCTGTTCTGCTGGAGGTGGC 3464
Db 2894 ProProPro-----LysValLeuThrSerProAlaThrThrProThrAlaThr 2909
Qy 3465 GAAGCCTCAATGACAACTGGGCTGACCCAGAACT----- 3500
Db 2910 SerSerLysAlaThrSerSerSerSerSerProArgThrAlaThrLeuProValLeuThr 2929
Qy 3501 -----ACAGCGCTACAGCAGCTCTCTGTACTTGGTGGCTGCTCAACCGTC 3548
Db 2930 SerThrAlaThrLysSerThrAlaThrSerValThrProIleProSerSerThrLeuGly 2949
Qy 3549 -----AGGCTGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3581
Db 2950 ThrThrGlyThrLeuProGluGlnThrThrThrProValAlaThrMetSerThrIleHis 2969
Qy 3582 CCCACTCTTGCACCCAGCTCTCCAGGC-----CTGCTGGGAGCATTCGGGCGTC 3631
Db 2970 ProSerSerThrProGluThrThrHisThrSerThrValLeuThrThrLysAlaThrThr 2989
Qy 3632 ATGGCAGCTCTGGGTCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3676
Db 2990 ArgAlaThrSerSerSerThrSerThrProSerSerThrProGlyThrThrTriPLeuThr 3009
Qy 3677 -----GAACCCACTTGG 3688
Db 3010 GluLeuThrThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSerSer 3029
Qy 3689 ACTCTGCGCTTGCACCCAGCTCTCCAGAGATGAGCAGCTTCTGCTGCTGCTGCTGCTG 3748
Db 3030 ThrProGlyThrThr-----TriPLeuThr 3038
Qy 3749 GAGCTGCAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3808
Db 3039 GluLeuThrThr-Thr-----AlaThrThrThrAlaSerThrGlySerThrAlaThrLe 3056
Qy 3809 CAGCTTCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3861
Db 3056 user-SerThr-----ProGlyThrThrTriPLeuThrGluProS 3070

QY 3862 --TGTACCTTCTCTCTCTCCCTTCAAAACCTGCTCTCCACCTTTGTGAGAGCCAG 3919
Db 3070 erThrThrAlaThrValThrAlaProProGlySerThrAlaThrAla-SerSerThrGln 3089
QY 3920 CCCTGTATGCCAACACTGTTGAGCCAGGACAGAACTGTGAGCCTCTGCGCCCTTCC 3979
Db 3090 AlaThrAlaGlyThrPro---HisValSerThrThrAlaThrThrProThrValThrSer 3108
QY 3980 TGGACCGCTGGGGCTGTGATGCGATCAGCCCTGCTCTCTCCCA---CCTCCCAAGGT 4036
Db 3109 -----SerLysAlaThrProSerSerProGlyThrAlaThrAla 3122
QY 4037 CTACCGAGTGGGAGGAGTACAGTACGCTGCTGCTGCTCTCTCTCTCTCTCTCTCT 4096
Db 3123 LeuProAlaLeuArgSerThrAlaThrThrProThrAlaThrSerPhe----- 3138
QY 4097 TGCTCGAGGAGTGTGAAGTGGTTCAGGTGGTTCAGAGGCGCTCATCGCTCTCTCTCT 4156
Db 3139 -----ThrAlaIleProSerSerSer 3145
QY 4157 TTGCTACCACTTGGCCAGTCCCAAGCCCTCAGTGGCAGCATCTGGAGGCGCAGG 4216
Db 3146 LeuGlyThrThrTrpThrArgLeuSerGlnThrThr 3157
QY 4217 GTTGAGGGCCACCCACACATGCTTTCTGGGGTGAAGCCCTTGGCTGCCCACTCT 4276
Db 3158 -----ThrProThrAla 3161
QY 4277 CCTTGGATGGTGTGCTCTCTATCCCAATCACTCTATACATCCAAATTCAGGAACA 4336
Db 3162 ThrMetSerThrAlaThrProSerSerThrProGluThrValHis-----Thr 3177
QY 4337 AACATGGTGCAATTTCTACACAAAGAGATGAGATTAAACAGTGGGGTGGGTCTGC 4396
Db 3178 SerThrVal----- 3180
QY 4397 ATTGGAGTCCCTATTAACAGAGAGAAATCTGAAGACACAGGGCA---GGGACA 4453
Db 3181 -----LeuThrThrAlaThrThrAlaHisThrThrLysValProThrThrThrGly 3196
QY 4454 GACCAGACAGACGAGGTCTCCAAAGCAGCAGAGTGGCAACAAACCCGAGCTGAGC 4513
Db 3197 ProSerSerThrProGlyThrAlaHisThrThrLysValProThrThrThrThrGly 3216
QY 4514 ATCAGACCTTGCTCGAATTTGTTCCAGTATTACGGTCTCTCTCTCTCTCTCTCTCT 4573
Db 3217 PheThrAlaThrProSer-----SerSerProGlyThrAlaLeuThrProPro--- 3232
QY 4574 CCAGGTATCTGGGTGGCCAGCTGGGAGGCGGCAACCATAGCCACACGACGATTC 4633
Db 3233 -----ValTrpIleSerThrThrThrProThrThrThrThrThrThrThrThr 3248
QY 4634 CTGAAAGTTTACAATGAGTAGCATTTTGGGTGTAGGGTGGCAGCTCCCAAGCCCTG 4693
Db 3249 -----SerGlySerThrValThrThrThrThrThrThrThrThrThrThr 3257
QY 4694 CCCCCAGCCCCACCACTCATGACTAAGTGTGTGTATTATATATTTATTTTGA 4753
Db 3258 IleProGlyThrThrHisThr----- 3264
QY 4754 GATGTTATTATAGATGATTATTATTCAGAAATTTCTATCTGTATTAAACAAATAAAA 4813
Db 3265 -----AlaArgValLeu 3268
QY 4814 TGCTTCCCCCAAGACTAGTCTCTTCTGCCCGCTC---ACCCCTCTGGTGCATCAG 4870
Db 3269 ThrThrThrThrThrValAlaThrGlySerMetAlaThrProSerSerThrThrGln 3288
QY 4871 ACTCTGCCACCCCTGGCTCCCACTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4930
Db 3289 ThrSerGlyThrProProSerLeuThrThrThrAlaThrThrThrThrThrThrThrThr 3308
QY 4931 GGGAGAGGCGCTCTCTCTCTCCCGCACTGGGGCGATGGGCGCAGCTTACCC--- 4987

Db 3309 ThrThrAsnProSerSerThrPro-----GlyThrThrProIleProPro 3323
QY 4988 -----ACTGCTGTGCTGCCACCAACCTTGTATCCCTCCACACAGCTTCT 5041
Db 3324 ValLeuThrSerMetAlaThrThr-----ProAlaAlaThrSerSerLysAlaThrSerSer 3342
QY 5042 GTCCACCCAGG 5053
Db 3343 SerSerProArg 3346
RESULT 8
CGH7L
collagen alpha 1(III) chain precursor - human
N:Alternate names: procollagen alpha 1(III) chain
C:Species: Homo sapiens (man)
C:Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000
C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511;
R:Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A:Reference number: S05272
A:Accession: S05272
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1240, 'V', 1242-1466 <PRC>
A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human procollagen
A:Reference number: S04642; MUID:89350838; PMID:2764886
A:Accession: S04642
A:Molecule type: mRNA
A:Residues: 1-1196 <ALA>
A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
R:Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
A:Reference number: PE0011; MUID:89378752; PMID:2777083
A:Accession: PE0011
A:Molecule type: DNA
A:Residues: 1-176 <BN>
A:Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814
R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 15, 7201, 1988
A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human
A:Reference number: S01726; MUID:88303360; PMID:3405773
A:Accession: S01726
A:Molecule type: mRNA
A:Residues: 1-170 <TOM>
A:Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061
R:Note: the authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen
A:Reference number: S04887; MUID:89386015; PMID:2780304
A:Accession: S04887
A:Molecule type: mRNA
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-616
A:Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045
R:Note: the authors' translation of residues 905-932 is inconsistent with the nucleot
R:Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide pep
A:Reference number: A90399; MUID:77134724; PMID:557335
A:Accession: A90399
A:Molecule type: protein
A:Residues: 'Y', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A:Experimental source: liver
A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-gal
R:Seyer, J.M.
submitted to the Atlas, December 1977

A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: V, 159-225, 229-277, A, 279-292, D, 294, S, 296-398 <SEY2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
splicing.
A;Reference number: I51868; MUID: 93304430; PMID: 8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 186-194 <MIL>
A;Cross-references: GB:S62925; NID: g386425; PIDN: AAD13937.1; PID: g4261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1
A;Reference number: S59511; MUID: 96067614; PMID: 7487954
A;Accession: S59511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID: g1195576; PIDN: AAB35615.1; PID: g1195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A;Reference number: A90414; MUID: 79000343; PMID: 687591
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675, N, 677-727 <SEY3>
A;Experimental source: liver
A;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A;Reference number: I55349; MUID: 91161621; PMID: 1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID: g180815; PIDN: AAA52041.1; PID: g180816
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
A;Reference number: A90438; MUID: 80198282; PMID: 6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895, A, 897-964 <SEY4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamanche, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A;Reference number: A38303; MUID: 91009133; PMID: 2145268
A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID: g180878; PIDN: AAB59383.1; PID: g
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
R;Mankoo, B.S.; Dalglish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID: 8819827; PMID: 3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, Y, 1020-1183, S, 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID: g300053; PIDN: CAA29886.1; PID: g300054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty
A;Reference number: A90446; MUID: 81208139; PMID: 7016180
A;Accession: A90446
A;Molecule type: protein

A;Residues: 965-979, A, 981-984, ES, 987, QN, 990-1096, P, 1098-1152, AT, 1155, S, 1155
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Bra
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procolla
A;Reference number: A93551; MUID: 85087944; PMID: 6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, P, 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID: g29584; PIDN: CAA25821.1
R;Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Remard, S.I.; Tolstoshev, P.; Bra
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the ty
A;Reference number: I52393; MUID: 86187804; PMID: 3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: GB:M13146; NID: g180415; PIDN: AAA52003.1; PID: g180416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long a
A;Reference number: I59025; MUID: 85216505; PMID: 3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID: g180417; PIDN: AAA52004.1; PID: g180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen
A;Reference number: A92516; MUID: 85157600; PMID: 2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, V, 1242-1356, P, 1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797;
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1411/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-D
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfid
er of their length, is formed with desmosine cross-links made from lysine and allyls
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains in
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; i
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willibrand factor type C repeat homology <WVC>
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;154-167/Region: amino-terminal nonhelical telopeptide
F;168-1196/Region: helical
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FC>
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predict
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predict
F;161,1212/Modified site: allylsine (Lys) #status predicted
F;263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F;584,1094/Modified site: carboxylate (Lys) (covalent) #status experimental
F;948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
F;1106/Binding site: carboxylate (Lys) (covalent) #status predicted

5059	QY	GGGAACCTGGGCTGCAGAAAGCTGTGTGGGAGGACTGAGGGATCAAGGGGTGTGGT	5060
216	Db	 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly	231
4999	QY	GGCAGCAGCAGTGGGTAAAGTCTGAGGTGCGCCCATCGCCCGAGTGGGGGAGGAAGAGGG	4940
232	Db	 GluSer-----GlyArgProGlyArgPro-----GlyGluArgGly	243
4939	QY	---CCTCTTCCAGAGCTCTGTGCAGCTCCACAGGCAAGCAGGAGGTG-----	4892
244	Db	 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe	259
4891	QY	---GGAGCAGGGGTGGCAAGAGTCTGATGAGCACACAGAGGGGTGAGGCTGGGCAAGA	4835
260	Db	 ProGlyMetLysGlyHiArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu---	277
4834	QY	GACTAAGTTCTGGGGCAAGCAATTTATTTGTTTAATACAAGATAGAAATTCGCAATAA	4775
277	Db	 -----	277
4774	QY	TATCATCTAATAATAACATCTCCAAATAATAATATTAAATACACACACTTAGAGTCA	4715
277	Db	-----	277
4714	QY	TGAGTGGTGGGCTGGGGGCGAGGCGCTTGGGAGCTGCCA-----	4673
278	Db	 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro	296
4672	QY	---CCCTACACCCAAAATGCTACTGCTATTGTAAACTTTTCAGGAATCTCTGTGGTGGC	4616
297	Db	 GlyProMetGlyProArg-----	302
4615	QY	TATGTTGCCCTCCCAGCCTGGCAACCCACAGATACCTCGGAAAGGGCGCAGAGAAGA	4556
303	Db	 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg---	320
4555	QY	GGCACCCTAATCTGGAAGACAAATTCGAGGCAAGGTCCTGATGCTCAGCTCGGGTTTTGT	4496
321	Db	 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro	335
4495	QY	TTGCCACTCTGTGCTTTGGAGACTCTCTGGTCTGGTCTGTCTGTCTCTGCCCTGTGCT	4436
336	Db	 GlyProProGlyThrAlaGlyPheProGly-----	345
4435	QY	TTCAGTATTCTCTCTCTGGTTTATATAGGGCACCTCCAAATGCAGACCCCAACCTGCAC	4376
346	Db	 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro-----	359
4375	QY	TTAATCTCATCTCTTTTGTGTAGAATTGCCACCATGTTTGTTCCTGAATTGGATGTAT	4316
359	Db	-----	359
4315	QY	AGAGTGATTTGGGGATAGGGAGCAACACCCATCCAAAGAGAGTGGGGCAGCAAAAGGC	4256
360	Db	 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro-----	371
4255	QY	TTCAACCCAGAAAGGCATGTGTGGTGGCGCCCTCAACCCCTGCCCTCCAGATGTGCCA	4196
372	Db	 ---GlyProGlnGlyHiAlaGly-----AlaGlnGlyProProGly-----	384
4195	QY	CCTGAGGGCTGGGTGGCCTGCGCAAGTGTAGCAAGAACAGCAGGAGGCCATGAGCGC	4136
385	Db	 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly-----	399

QY 1234 GGGATGGAGGAGGGGAAACACTCTCTGCAAGACGGGAGAGAAAGAGGCGATGTGTATG 1175
||||| |||
Db 1031 GlyAspArgGly----- 1034

QY 1174 TTTGAGGATGAGGACTTTGGCAGCGCATGAGTGTCTATCAGTAGATACATTGAGAATGGATCT 1115
||||| |||
Db 1035 -----GluAsnGlySer 1038

QY 1114 GTGTGTGCTGAGAGAGTCTTTGGCATGTGTGCCACTGTGTGTGTCACCATGTATAT 1055
||||| |||
Db 1038 ----- 1038

QY 1054 GTGTGTACATGAGGAGGAGCGGGCAGGCTAGGACTGCAGGTGCAGCTGCTGTGTTC 995
||||| |||
Db 1039 -----ProGlyAla---Pro-GlyAlaProGlyHisProGlyProProGly 1052

QY 994 AGTTGTA-----CCCTAGCTTCTCC 974
||||| |||
Db 1052 yProValGlyProSerGlyLysSerGlyAspArgGlyGluThrGlyProAlaGlyProSe 1072

QY 973 CAGTCTTCAGTCCCACTGCAGCTTGGGGAGTCACAGGTGGGAGACACACATGTCA 914
||||| |||
Db 1072 rGlyAlaProGlyPro-AlaGlyAlaArgGlyAlaProGlyProGlnGlyProArgGlyA 1092

QY 913 GAGCTGGCAGGCTGAGCACAGTCAGCTCTAAGGGAGGAAGCAGGAGAGAGCGCGGCCA 854
||||| |||
Db 1092 splysGlyGluThrGlyGluArgGlySerAsnGly-1lelvsGlyHisArgGlyPhePro 1111

QY 853 AGAACAGATCACCCATAGGAAGCAAGGTATAGGTGTAGACCATAGGCTGGGCTGCCT 794
||||| |||
Db 1112 GlyAsnProGlyProProGlySerPro-----GlyAlaAlaGlyHisGlnGly- 1127

QY 793 TATGGGCTTAGTTTGGGTATATCTTAGGATGTGATCTCTGAGTGTGGAGGTAGGG 734
||||| |||
Db 1128 -----AlaIleGlySerProGly 1133

QY 733 TGGGGTTAGATTGAAGGAGATTTAGTCTCACTCTCAGGGAACGGTGGCAGTGTGGA--- 677
||||| |||
Db 1134 -----ProAlaGlyProArgGlyProValGlyPro 1143

QY 676 ---GGTTCAACCGTCTTGGTGTAGTGGGGAAGATGTGAGCTCAATGAGGCGCTTTC 620
||||| |||
Db 1144 HisGlyProPro-----GlyLysAspGlyThrSer- 1153

QY 619 CATCAGCTGCAAGGTGGGACAGGAGAGAGAGTGCATTTGACGAGTCCGCGCTTG 560
||||| |||
Db 1154 GlyHisProGlyProIleGlyProProGlyProArgGlyAsnArgGlyGluArgGly--- 1172

QY 559 TTGGCTGATGGCACTGCAGAGTTTGGACATCCCAAGGATACCGAGAGTCCAGAGCGCAGGG 500
||||| |||
Db 1173 ---SerGluGlySerPro-----GlyHisProGlyGlnProGlyProProGlyProPro 1189

QY 499 TTGGGTGCTGTAGGAGGATGTGGGACAGTGGGGAATATGCTTTTCTGGATTTTCAGT 440
||||| |||
Db 1190 ---GlyAlaProGlyProCysCysGlyGlyGlyAlaAlaAlaIleAlaGlyVaiGlyGly 1208

QY 439 GAGGACCAAGGTGAGGATTCAGTGGACATTTGGA 405
||||| |||
Db 1209 GluLysSerGly--GlyPheSerProTyrThrGly 1219

RESULT 10
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Rifov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2187 <YOT>

A:Cross-references: EMBL:U48363; NID:gl666688; PID:gl666689; PIDN:AAB18732.1

C:Genetics:

A:Gene: Naca

A:Map position: 10

A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3

A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding a
C:Keywords: alternative splicing; DNA binding; transcription factor

Alignment Scores:

Pred. No.: 3.3e-12 Length: 2187
Score: 323.00 Matches: 435
Percent Similarity: 31.11% Conservative: 194
Best Local Similarity: 21.51% Mismatches: 762
Query Match: 3.42% Indels: 631
DB: 2 Gaps: 89

US-09-931-704-3 (1-5087) x T30826 (1-2187)

QY 60 CTCTCCACTCCGCGAGCTCCGGGAGAGGCGCGACCC----- 98

Db 105 PropProIleSerProAlaAlaLeuAlaLeuAlaSerProMetIleGlyLeuAlaGlnLys 124

QY 99 --GGCGCGCCAGCCAGCCCGCATGGACCTCCGAGCAGAGTTGAAACCCCAACTAGCC 155

Db 125 GlyAlaArgSerSerAlaProLeuSerLeuValAla----- 137

QY 156 CTGCTCTTCTATAACATGACAGCAGCGCCCATCTGATA----- 194

Db 138 -----LeuAlaProHisSerValGlnLysSerValCysPro 150

QY 195 CCTAACCCGACCAAGTCACAGCCCTCCAACTCACCCTCGCTGCGCCAGCAGCTCACCACA 254

Db 151 ProHisProLeuThrSerProSerAlaAlaGlyAlaGluLeuGlyAlaLeuThrAla 170

QY 255 TCCTTGTGGCTCAAAACCTCAACCGCCTAAATCAACCAAAATCCCAAGCTAAACTAATC 314

Db 171 SerIle----- 172

QY 315 TGAAACTTTTAAAGTAACCCAGTCTCTTAAACCTAACCTAGAGCCCAATGATATATCTA 374

Db 173 -----ProProLeuGluPro----- 177

QY 375 CCCTAGCCAAACCCCTAACTGCTTTGCCAGTCCAAAGTGTCCACTGAATCCTCACCTGG 434

Db 178 -----LysThrSerThrSerGlnValPro--- 185

QY 435 TCCTCACTGAAATCCAGAAAAGCATATTTCCCACTGCCACATCCCTCTTACAGCA 494

Db 186 SerGlnGlyThrLeuAsnLeuLysGlyThrAlaProCysProProAspValValArgAla 205

QY 495 -----CCCAACCCCT---GGCCTCTGGACTCTCTGGTATCTCTGGGATGTCCA 536

Db 206 PheProSerHisLeuGluAsnProLeuAlaSerValGlnProGlyLeuMetSerCysPro 225

QY 537 AACTCTGAGTGCCATCAGCCAAAGCCGACTCGTCAATGCACCTCTCTCCCTTCTCT 596

Db 226 GlnThrLeu-----Ser-AsnThrSerProValLysGlyValProIleSerSerAlaLe 243

QY 597 GTCCCCACCTTTCAGGCTGATGGAAAGGCTCTATTGAAGTCCCACTTTTCCCACTTA 656

Db 243 uThr-----GlnSerArgLeuSerLeuAsnLeuLysGlyProValSerProPro-- 259

QY 657 CACCAAGAACGGGTGAACCTCCACACTGCCACCGCTTCTCTGAGAGTGAGCTAAATCT 716

Db 260 -----AlaArgAsnThrAlaAlaProSerIleProLeuAlaProSerThrSerLe 276

QY 717 CTTTCAATCTAAACCCACCTTACACTTCCACACTCAGGAATCAGATCTCTAGATATACC 776

Db 276 uGlyCysHisLeuProLeuLeuHis-----HisSerSerValAspSerProIleGlnPr 294

Db 901 LysThrProLysSerValSerLeuLysGlyAlaProAlaMetThrSerLysLys----- 918
Qy 2631 TCTGGTTCCAAATGGGCCACTCTCATCTCTCTGTTGTTCCITGTCAGAAAACCTTTGCTT 2690
Db 918 ----- 918
Qy 2691 CACTCCACTGCCCTCTCTAGTTCCGACCCCTTTTCTCTCTGCTTTTCCCTGCCAAATT 2750
Db 919 ---AlaThrGluIleAlaAlaSerLysAspValSerProSerGlnPheProLysGluVal 937
Qy 2751 -----TCTCCA-----AGG 2759
Db 938 ProLeuLeuGlnHisValProThrSerProProLysSerProValSerAspThrLeu 957
Qy 2760 AGTGGTCTACACCTCTGCTCTCA-----CTTCTCT 2789
Db 958 SerGlyAlaLeuThrSerProProProLysGlyProProAlaThrLeuAlaGluThrPro 977
Qy 2790 CTCACCCACTCTCTTTAAACCCCTGCAATCTGGCTTCAGGCCCCAGCAATGGTTCT 2849
Db 978 ThrTyProLysLysSerProLysProAlaAlaSerLysLysThrProAlaThrProSer 997
Qy 2850 CTCCAAGGTCGTGAGGACCTCTCTGCAAGCCGACAGTGT----- 2894
Db 998 ProGluGlyValThrAlaValProLeuGluIleProProCysSerLysLysAlaProLys 1017
Qy 2895 -----GAAGGCTCATCTCTCTGCTGCTGTTTTCAGCCACACTGCTGA 2939
Db 1018 ThrAlaAlaProLysGluSerAlaThrSerSerLysArgAlaProLysThrAla 1037
Qy 2940 GGCCTG-----CTGCTCTCTCA-----ACTCCTCTCTGCTGCTCTGCTCTCTCT 2987
Db 1038 ValSerLysGluIleProSerLysGlyValThrAlaValProLeuGluIleSerLeuPro 1057
Qy 2988 GGGCCACCTCTACCTCTCCAGCTCTCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3047
Db 1058 LeuLysGluThrSerLysSerAlaThrProGlyGluLysSerAlaSerSer----- 1074
Qy 3048 AGCGGGCACTCTCCCAAGGTTTGGCCACCCAGCAATCAGCACGCTCTCTCTGAGCGTCT 3107
Db 1075 ProLysArgSerProLysThrAlaGlyProLysGluThr-----ProProGlyGlyVal 1092
Qy 3108 TGTGCTCTCTCTCT-----CTCTCTTTTCTACGCTCTCTCCATTTGGAGAGCT 3155
Db 1093 ThrAlaValProProGluIleSerLeuProProLysGluThrProGln----- 1108
Qy 3156 CACCACCGCACTGCTCACTGCTACCTGCATACAAATGATATCTCTATTTGNAAACT 3215
Db 1109 -----AsnAlaThrProAsnGluSerLeuAlaAlaSerSer 1120
Qy 3216 CAG-----GGAGGC----- 3224
Db 1121 GlnLysArgSerProLysThrSerValProLysGluThrProProGlyGlyValThrAla 1140
Qy 3225 -----CATGAACAAGAGCTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3278
Db 1141 MetProLeuGluIleProSerAlaProGlnLysAlaProLysThrAlaValProLysGln 1160
Qy 3279 ATAGAAACTTTGGGAGCAGTATCTCTGTTGGTGGAGCAGCGGCTCTGCTCTCTCTCTCT 3332
Db 1161 IleProThrProGluAspAlaValThrIleLeuAlaGlySerProLeuSerProLysLys 1180
Qy 3333 -----CT 3344
Db 1181 AlaSerLysThrAlaAlaProLysGluAlaProAlaThrProSerValGlyValIleAla 1200
Qy 3345 -----TCACCTCTCTCTTTTCACAGCTGAATCTACCTGGGCGCCCTTTTCA 3389
Db 1201 ValSerGlyGluIleSerProSerProLysLysThrSerLysThrAlaAlaProLysGlu 1220
Qy 3390 ACGAGCCAGACTTCAACCTCTCCCGCTGGGGCAGAGACTCTGCCCAGCGGCGCTGTTG 3449
Db 1221 AsnSer-----AlaThrLeuProProLysArgSerProLysThrAlaAlaProLysGlu 1238

Qy 3450 ACTTGGAGGTGTGGGAGAGCCTCAATGACAAACTGCGGCTGACCCAGAACTACGAGSCCT 3509
Db 1239 ThrProAlaThrSerSerGluGlyValThrAlaVal-----ProSerGluIleSerPro 1256
Qy 3510 ACAGCCACTCTTCTGTGTTACTTTCGCTGGCTCAACCGTCAGGCTGCCACTCTGCTGAGCTGC 3569
Db 1257 SerProProThr-----ProAlaSerLysGlyValProValThrLeuTh 1271
Qy 3570 GCGCAG-----CCTGGCCCACTTCTGCACGACCTCCAGGCT----- 3609
Db 1271 rProLysGlyAlaProAsnAlaLeuAlaGluSerProAlaSerProLysLysValProly 1291
Qy 3610 -----GCTGGGCGAGCATTTG 3623
Db 1291 sThrAlaAlaProGluGluThrSerThrThrProSerProGlnLysIleProLysValAl 1311
Qy 3624 CGGGCGTCATGCAGCTCTGGGCTACCC----- 3651
Db 1311 aGlyProLysGluAlaSerAlaThrProProSerLysLysThrProLysThrAlaValPr 1331
Qy 3652 -----ACTGCCCCAGCGCTGCTGGAGCTGAACCCACTTGGACTCTCTGCGCTG 3701
Db 1331 oLysGluThrSerAlaProSerGluGlyValThrAlaValProLeuGluIleProProse 1351
Qy 3702 CCCACAGTGACTTCTCT-----CCAGAAGATGGACGACTTCTGCTGCTGAAGG 3749
Db 1351 rProArgLysAlaProLysThrAlaAlaProLysGluThrProAlaProSerProGluGl 1371
Qy 3750 AGCTCAGACCTGCTGCGCTCGGCCAAGGACTTCAACCGGCTCAAGAAGAAGATGC 3809
Db 1371 yAlaThrThrAlaProValGlnIleProProSerProArgLysGlySerLysLysAlaGl 1391
Qy 3810 AGCTCCAGCAGCTGACCTACCTGACCTGACCTGCTCTCTCTCTCTCTCTCTCTCTCT 3846
Db 1391 ySerLysGluThrProThrThrProSerProGluGlyValThrAlaAlaProLeuGluI 1411
Qy 3847 -----TCATGGCTCTGACCTTCTGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 3884
Db 1411 eProLysSerLysLysThrSerLysMetAlaSerProLysGluThrLeuValThrPr 1431
Qy 3885 CCCTTCA-----AACCT 3920
Db 1431 oSerSerLysLysLeuSerGlnThrValGlyProLysGluThrSerLeuGluGlyAla 1451
Qy 3921 CCTGTATGCCAACCTCTTCTGAGCA-----GGAGACAGAGCTGTGAGCTCTC 3968
Db 1451 hrAlaValProLeuGluIleProProSerHisLysLysAlaProLysThrValAspPro 1471
Qy 3969 TGGCCCTTCTCTGACCGCTGGGCTGTGATGCG-----A 4004
Db 1471 ysGlnValProLeuThrProSerProLysAspAlaProThrThrLeuAlaGluSerPro 1491
Qy 4005 TCAGCCT 4064
Db 1491 erSerProLysLysAlaProLysThrAlaAlaProProSerGluArgValThrThrValp 1511
Qy 4065 GCCCTGCT 4124
Db 1511 roProGluLysProAlaThrProGlnLysAlaSerGlyThrThrAlaSerLysValPro 1531
Qy 4125 TTGGTGCA-----GAGGCGCTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4175
Db 1531 alProAlaGluThrGlnGluValAlaValSerSerArgGluThrProValThrProAlaV 1551
Qy 4176 TGCCACCCAGCCCTCAGTGGGACATCTGGAGGCGAGGGGTGAGGGCCACCCACAC 4235
Db 1551 alProPro-----VallysaenProSerSerH 1560
Qy 4236 ACATGCTTTCTGGGGTGAACCCCTTTGGCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4295
Db 1560 isLysLys-----ThrSerLysThrIleGluLeuLysGluAlaPro-----AlaT 1575

Db	283	GlyGlnLysGlyGluLysGlyGlnGlnGluProGlyLysArgGlyLysProGlyLys	302
Qy	1338	GCTCATGCTTCTCGTCCATCTCTCCACAGGGGACTCGTGGGGATGTTAGCGTGCTGT	1397
Db	303	AspGlyGluAenGly-GlnProGlyLeuProGlyLeuProGlyAspProGlyTyrPro--	321
Qy	1398	GCAGGTGCTCTGGCACCTCCCTGCAGTGCACGCTCTCAATCGCACAGGGACCCAGGC	1457
Db	322	-----GlyGluProGlyArgAspGlyGluLysGlyGlnLysGlyAspThrG1	337
Qy	1458	CTGGCCCTCCATCCAGAAAACCTATGACCTCACCCG	1494
Db	337	yProGlyProGlyLeuValleProArgProGlyThrGlyleThrleGlyG1	357
Qy	1495	-----CTACCTGGAGCACCACTCCGACGTTGGCTGGGACCTATGTGA	1538
Db	357	uLysGlyAsnleGlyLeuProGlyLeuProGly-	368
Qy	1539	GTATCCAGCGTAGGAATCTGGAGTTGGGA-----GGAGTGAGAGT-	1581
Db	369	-----GluLysGlyGluArgGlyPheProGlylleGlnGlyProProGlyLe	384
Qy	1582	-----TGGGAAAGACAGTCTTAACCGTGGAGGTTCTGGTAAATGATGGGGTGA	1631
Db	384	uProGlyProGlyAlaAlaValMetGlyProProGlyProProGlyPheProGlyG1	404
Qy	1632	GGAGGGCTCTTGGCTCCACGACGCTCCCTCTGTGCTATCTCTGCCCTTCCCTCT	1691
Db	404	u-ArgGlyGlnLysGlyAspGluGlyProProGlylleSerleProGlyProProGlyL	424
Qy	1692	TAGTGGCCCCCCTATCCATCCCTG-----GCCACGAGACTAGGCATGGGCAG	1745
Db	424	euAspGlyGlnProGlyAlaProGlyLeuProGlyProProGly-	438
Qy	1746	GCCTCGCACCCGCTTGGCCCATTCGCCCATGGC-----TCCAGCCCGACCGCC	1796
Db	439	-----ProAlaGlyProHisleProProSerAspGluLecysGluProGlyProP	456
Qy	1797	CGCTCCCGCTGGGGCGGGGAAGTCTCTCTGTTTACCCGTGTGTGGTGTCTCTTG	1856
Db	456	roGlyProProGlySerProGlyAsp-----	464
Qy	1857	CGCGGGCGGGTGTGGTGGGACAGAGG-----	1885
Db	465	-----LysGlyLeuGlnGlyGlnGlnGlyValLysGlyAspLysGlyAspThrCysPheA	483
Qy	1886	-----GCCACCTCCCATGCCTGGTCCAGCTCGCTCTG	1922
Db	483	snCysleGlyThrGlylleSerGlyProProGlyGlnProGlyLeu-	498
Qy	1923	CCCCAGACCTGGGCGCTGTGCTCTGGACCCAGGGGCTCCCTTCGCTCGCTCTCC	1982
Db	499	-----ProGly-----LeuProGlyProProGlySerLeuGlyPheP	511
Qy	1983	CATCTTAGCTGGGCTCTTAGGG-----GGGTCTATGGGG-----AAGGGGACTGTAGGGA	2033
Db	511	roGlyGlnLysGlyGlnLysGlyGlnAlaGlyAlaThrGlyProLysGlyLeuProGlyI	531
Qy	2034	ACCAGGACGATGCGCAGG---GGGTTAGGTGTGGATGAGGTTATGCTCTAAGGAT	2090
Db	531	leProGlyAlaProGlyAlaProGlyPhePro-----Glys	543
Qy	2091	TTGGGGGTGGTCCA---CAGGTGTTTCAGAGACCCAGGAGAGAAAGAGAGG-	2141
Db	543	erLysGlyGluProGlyAspIleLeuThrPheProGlyMetLysGlyAspLysGlyGluL	563
Qy	2142	TTGAGGAGCGCAGGCA-----CCATGGGGAAACCGGCCCTCTTCC	2183
Db	563	euGlySerProGlyAlaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly	582
Qy	2184	CGTGTCTCTTCACATCCAGACCCCTACTCTCGAGCCAGGGAAGAAAGAGAGAGAG	2243
Db	583	LeuProGlyLeuProGlyProLysGlyGlnProGlyGlyleThrPheLysGlyGluArg	602

[illegible]

Db	543	erLysGlyGluProGlyAspIleuThrProGlyMetLysGlyAspLysGlyGlu	563
Qy	2142	TTGATGGAGCGAGGCA-----CCATGGGGAAACGGGCCCCCTCTTC	2183
		:::	
Db	563	euGlySerProGlyAlaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly	582
Qy	2184	CGTGTTCCTTCCACATCCAGACCCCTACTCTCGAGCCAGGGAAGAAAGAGAGAG	2243
		:::	
Db	583	LeuProGlyLeuProGlyProLysGlyGluProGlyGlyIleThrPheLysGlyGluArg	602

[illegible]

Db	563	euGlySerProGlyAlaIaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly	582
QY	2184	CGTGTTCCTCTCCACATCCAGACCCCTACTCTGGAGCCAGGAAAGAGGAGAGAG	2243
Db	583	LeuProGlyLeuProGlyProGlyGlyGluProGlyGlyIleThrPhenylsGlyGluAsp	602

[illegible]

D_b 583 LeuProGlyLeuProGlyProLysGlyGluProGlyGlyIleThrPheLysGlyGluArg 602

Db 1443 -----LeuAspGlyProProGlyProAspGlyLeuGlnGlyPro----- 1455

QY 5042 GTCCACCCAGGTTTCCT 5060
||||| |||

Db 1456 -----ProGlyProPro 1459

RESULT 12

G86292
hypothetical protein F7H2.17 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
R;Accession: G86292
P;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White,
Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maifi, R.
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: G86292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1006 <STO>
A;Cross-references: GB:AE005172; NID:G8927662; PIDN:AAF82153.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Alignment Scores:

Pred. No.:	5,17e-12	Length:	1006
Score:	319.50	Matches:	247
Percent Similarity:	27.35%	Conservative:	64
Best Local Similarity:	21.72%	Mismatches:	310
Query Match:	3.39%	Indels:	517
DB:	2	Gaps:	50

US-09-931-704-3 (1-5087) x G86292 (1-1006)

QY 2151 CCGAGGCACCATGGGAACCGGCCCTCTTCCCGTTCCTCTTCCA----- 2198
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 103 ProArgProArgProArgProArgProArgProArgProArgProArgProValProSer 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2199 -----CATCCACACCTACTCTGGAGCCAGGAACAAGAAGAAAG 2243
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 123 ProProProProLeuHisProArgPro----- 131

QY 2244 GTGGCGGGGAGCTGGCTCCAGCCCCAGGATACACCGAGAAATTAGTTGTCTCTGTGC 2303
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 132 -----SerProCysProProProLeuMetProSerProProProLeuVal 146

QY 2304 TTGTACGCGTGAACTCCCTCGGGCCCTTG-----CCTATCCAGCCCTCTCC 2354
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 147 ProSerProProProProProSerProSerProValProSerProProProSerPro 166

QY 2355 CCTTGTCTTCCTTCTTTCCAGTTATACATCCCTCATCCCTTTCCT----- 2405
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 167 ProProPhePhePheProSerProProProProValIleValPheProProLeu 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2406 -----GGGCCCCAGCGCTCCCGGAGGGTTCGAAAGGCTCTGCCCTTCCCTTA 2456
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 187 ValProSerProProProProProGlyAspGlnThrGlnProProProLeu 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2457 TACCATGCTGTCTTCCATAGCCCTTCTCTGTCTCTACTCATGAGCTGCCATTTCTT 2516
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 207 Trp-----LeuProPro-----ProProPheGly 214
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2517 CTTTCTGAACCTGCTTCCTATCAGCTGAACCTTCTTTCGGAGTGTAGTAGTACCG 2576
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 215 AspGluThrProValPheSerLeuProPro-----ProLeuAspGluPro 231
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 709 -----AlaProGlyIleThrGlySerProProProLeuVal 721
 QY 4481 AGC-----ACAGAGTGGCAACAAACACCGAGCTGAGCATCAGG 4519
 Db 722 AlaAspValProProMetProProLeuAlaTrpPheSerProProAspIleThr----- 739
 QY 4520 ACCTTGCTCGAATGTCTTCCAGTATTACGGTGCCTTCTCTGCCCTTTCOCAGGG 4579
 Db 740 -----ThrGlySerProProProVal 748
 QY 4580 TATCTGGGTGGCAGGCTGGGAGGCAACATA-----GCCACACACAGAT 4630
 Db 749 PheLeuLeuProProProLeuAspArgSerThrLeuThrProProAlaAlaProValAsp 768
 QY 4631 TTCCTGAAAGTTTACAATGACGTAGCATTTTGGGGTGTAGGGTGGCAGCTCCCAAGGCC 4690
 Db 769 -----AsnLeuProProVal 773
 QY 4691 CTG-----CCCCCAGCCGCCACCCACTCATGACTCTTAAGTGTGTGTTAATAATT 4741
 Db 774 IleIleThrGlySerProProProValAsnAsn----- 784
 QY 4742 TATTTATGGAGATGTTATTATTAGATGATATTATTGCGAATTTCTATTCTGTAT 4801
 Db 784 ----- 784
 QY 4802 TAACAATAAAATGCTTGGCCCCAGAACTTAGTC----- 4834
 Db 785 -----LeuProProAspIleValIleGlyGlnProProProAspVal 799
 QY 4835 -----TCTTGGCCAGGCTCACCCCTCTGGTGTCTCATCAGACTT 4876
 Db 800 ThrIleGluProProIleAspGlnSerThrLeuThrProProValIleProValThrLe 819
 QY 4877 GCCACCCCTGGCTCCCACTCCCTGCTTGGCTGTGGTGGAGCTGCACAGAGCTCTGGGAG 4936
 Db 819 uProPro-----ProValGlnAspLeuProSerIle----- 829
 QY 4937 AGGCCCTCTCTCCCGCAGCTGGGGGATGGGGCGACCT---CAGACTTACCCACTGCT 4993
 Db 830 -----LeuProProAlaAspGluLeuProProProValGlnPheProIle 847
 QY 4994 GCTGCCACACCAACC-----CCTTGATCCCTCAGTCTCCCAACACAGCTTCTGTC 5044
 Db 847 eLeuProProProValGlnAspPheProProIleLeuAlaProProAlaAspGluPhe-P 867
 QY 5045 CACCCAGGTTTCCCTCACCCACCTTGTGCTAGCTTCTCT 5085
 Db 867 roPro-----AsnLeuProProValLeuGluPhePro 878

RESULT 13

S06733
 hydroxyproline-rich glycoprotein precursor - common tobacco
 C:Species: Nicotiana tabacum (common tobacco)
 C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
 C:Accession: S06733
 R:Keller, B.; Lamb, C.J.
 Genes Dev. 3, 1639-1646, 1989
 A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene
 A:Reference number: S06733, MUID:90128263, PMID:2612909
 A:Accession: S06733
 A:Molecule type: DNA
 A:Residues: 1-620 <KEL>
 A:Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867
 C:Superfamily: hydroxyproline-rich glycoprotein
 C:Keywords: glycoprotein

Alignment Scores:

Pred. No.:	1,02e-11	Length:	620
Score:	314.50	Matches:	217
Percent Similarity:	30.36%	Conservative:	42
Best Local Similarity:	25.44%	Mismatches:	268
Query Match:	3.33%	Indels:	326

Db: 2 Gaps: 44
 US-09-931-704-3 (1-5087) x S06733 (1-620)
 QY 1658 CCCCTGTCTGTCTATCTCTCTGCTTCCCTCTTAGT---GGCCCCCCTTCCCA 1714
 Db 36 ProProValThrSerGlnProProSerSerIleGlyLeuSerProSerAlaPro 55
 QY 1715 TCCCTGGCCCGAGGACTAGGATGTGGGAGGCTCGACCCCTTGGCCCTTCCCT 1774
 Db 56 ThrThrThrProProSer-----ArgGlyHisValProSerProArgHisAlaPro 72
 QY 1775 ACTGCTGCGACCCAGCCGCCCTCCCTCTGGGGCGGGGAAAGTCTCTCTGTTA 1834
 Db 73 -----ProArgHisAlaTyProProProSerHisGlyHisLeu----- 85
 QY 1835 CACCTGTGTGGTGTCTCTTCCGGGGGGGGTGGTGGG----- 1876
 Db 86 -----ProProSerValGlyGlyProProProHisArgGly 97
 QY 1877 -----GACAGAGGG---GCCCCACCTCCCATCGCTCGTCCAGC---TCGCC 1918
 Db 98 HisLeuProProSerArgGlyPheAsnProProProSerProValIle-SerProSerHi 117
 QY 1919 TCTGCCCGCAGACCTGGGG-----CCCTGTCTCTGACCCAGGGGCTCCCTTCCGT 1972
 Db 117 sProProProSerTyGlyAlaProProProSerHisGlyProGlyHisLeuProSer-- 136
 QY 1973 CTGCTCTCCCATCTCTAGTGGGCTCTAGGGGGTCTATGGGGGAAAGGAGCTGTAGG 2032
 Db 137 -----HisGlyGlnArg----- 140
 QY 2033 AACCCAGGAGTAGTGGCAGGGGGTTAGGTGTGATGAGGTATTCTGTAAAGATT 2092
 Db 140 ----- 140
 QY 2093 GGGGTGTCTCAGAGGTGTTCAGAGCCAGGAGAG-----AAGGAAGAGG 2140
 Db 141 -----ProProSerProSerHisGlyHisAlaProProSerGlyG 154
 QY 2141 GTTGAGGAGCGGAGGACCATGGGAAACCGCCCTCTTCCGTTGTTCTCTTCA-- 2198
 Db 154 yHisThrProProArg-----GlyGlnHisProProSerHisArgArgProSerProPr 172
 QY 2199 -----CATCCAGACCTCTCTCTGGAGCCAGGAAAGAAAGGAAGGAGTGGCGG 2251
 Db 172 oSerArgHisGlyHisPro----- 178
 QY 2252 GGAGCTGCTCCAGCCCGCAGGATACACCGAGGAAATAGTTTCTCTGTCTGTTCAGC 2311
 Db 179 -----ProProProThrTy----- 183
 QY 2312 GTGTGAACCTCCCTGGGCGCTTG-----CCTATCCAGCGCTCTCC-----CCTTG 2359
 Db 184 -AlaGlnProProProThrProIleTySerProSerProGlnValGlnProProTh 203
 QY 2360 CTCTCCCTCTTCTTCCAGTTATACATCTCCCTCATCCCTTCCCT-----GGGCC 2410
 Db 203 rTySerProProProProThrHisValGlnProProProProSerProSerArgGlyHi 223
 QY 2411 CCAGCGCTCCCGAGGGTTGGAAAGGCTCTGCCCTCTTCCCTATATACCTGTCTT 2470
 Db 223 sGlnProGlnProProThrHisArgHisAlaProProThrHisArgHisAlaProProTh 243
 QY 2471 CCATAGCTTCTCTGCTCTCTACTCATGAGACTGCTCCATTTCTCTCTCTGCAACCT 2530
 Db 243 rHisGlnProSerPro-----LeuArgHisLeuProPro----- 254
 QY 2531 GCTCTATCAGCTGAACCTTCTTTCGGAGGTGTAGTGAGTACCGCTCTCTCCCGAGCC 2590
 Db 255 -----SerProArgArgGlnProGlnProProThrTy 265
 QY 2591 CTCAGCTGGTGGGCTGGGTGTGTGTCAGCGCAATGGGGCTCTGTTCATGGGCCACT 2650

```
Db 265 rSerProProProAlaIaIaGlnSerProGlnProSerProThrTyrSerPro-- 284
QY 2651 CTATCTCTCTCTGTTCTGTCAGAGAAACCTTTGTTCTTCACTCCACTGCGCTCTAG 2710
Db 285 -----ProProProThrTyrSerProProProProSe 295
QY 2711 TTCCCGACCCCTTTTCTCTCTGCTGCTTCCCTGCGCAAAATTTCTCCAGGAGTGCTACA 2770
Db 295 r-----ProIleTyrSerPro-----ProProAlaIaTyrSe 306
QY 2771 CCCTCTGCTCACTTCTCTCCACCACTCACTTCTTAAACCCCTGCAATCTGGCTTCC 2830
Db 306 rProSerProProProThrProThrPheSerProProProProAlaIaTyrSerPr 326
QY 2831 AGGCCCGAGCAATGGTCTCTCCCAAGTCTGAGGCGTCTGAGGCGTCTGAGGCGT 2890
Db 326 oProProThr-----TyrSerProPro----- 333
QY 2891 TTTTGAAGGCTCATTTCTCTGCTGCTGTTTGTGAGCGACACTGCTGAGCGCTGCTGCC 2950
Db 334 -----ProPr 335
QY 2951 TTCTCGAACTCTCTCTCTGCTGCTCTCTGCTGCTCTCTGCTGCTCTCTCTCTCTCT 3007
Db 335 oThrTyrLeuProLeuProSerProIleTyrSerProProProProValTyrSerPr 355
QY 3008 AGCTCTCTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3067
Db 355 oProProProProSer-----TyrSerProProProProThrTy 368
QY 3068 TTGCCCAACCCAGCAATCAGCAGCTCTCTCTGAGCGTCTGCTGCTCTCTCTCTCTCT 3127
Db 368 rLeuProProProProProSerPro-----ProProProSe 381
QY 3128 TTTTCTCAGGCTCTCTCA-----TTGAGAGCTCACCACCGCCACTGCTTCACTGCTCA 3181
Db 381 rPheSerProPro--ProProThrTyrGluGlnSerProProProProAlaIaTyrSer 400
QY 3182 CCTGCATCAATGATATCTCTTATTTGAAAACTCAGGAGGCGCATGACAAAGAGCT 3241
Db 401 ProProLeuProAlaProProProThrTyr-----SerProProProProTyrSerPro 418
QY 3242 AGCATGGAGACA-----GGGCGAGTGTGAGCGGCTCAGGCGGACACAAAAATAGAACTTT 3289
Db 419 ProProProThrTyrAlaGlnProProProLeu----- 429
QY 3290 GGGAGCAGGTATCTCTTGTGTTGAGCGGCTCTGCGCTCTCTCTCTCTCTCTCTCTCT 3340
Db 430 -----ProProThr--TyrSerProProProProAlaIaTyrSe 441
QY 3341 CCCATCACCTCTCTCTTTTTCACAGCTGAACTACCTGGGCGCCCTTTTCAACGAGCCAGAC 3400
Db 441 rProProProProPro-----ThrTyrSerProProProProProThrTySe 456
QY 3401 TTCAACCTCCCGCTGCGGCGCAGAGACTCTGCGCGGCGCCTGTTGACTTGGAGGTG 3460
Db 456 rProProProProAlaIaTyrAlaGlnProProProProPro----- 470
QY 3461 TGGCGAAGCCTCAATGACAACTGCGGTGACCCAGAACTACGAGGCTTACAGCCACTTT 3520
Db 471 -----ThrTyrSerProProProProProAlaIaTyrSerPro--ProProProSe 485
QY 3521 CTGTGTTACTTCTGCGGCTCAACCGTCAAGGCTGCCACTGCTGAGCTGGCGCGCAGCTG 3580
Db 485 erProIleTyrSerProProProProGln-----ValGlnProL 498
QY 3581 GCCCATCTCTGACACAGGCTCCAGGCGCTGCTGGGCGAGCATTCGGGGCTCATGGCAGCT 3640
Db 498 euProProThrPheSerProProProProArgIleHis----- 511
QY 3641 CTGGGCTACCCACTGCCCC-----CAGCGCTGCTCTGGAGCT----- 3676
```

```
Db 512 -----LeuProPro--ProProHisArgGlnProArgProProThrProThrTyrGlyGln 529
QY 3677 -----GAACCCACTGAGCTCTCTGCGCCT-----GCCACAGTGAAGTCTCTC 3718
Db 530 ProProSerProProThrPheSerProProProProArgGlnIleHisSer--ProProPr 549
QY 3719 CAGAAGATGACAGCTTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3778
Db 549 oProHisTyrGlnPro-----ArgThrProThrTyrGlyGlnProPr 565
QY 3779 AAGACTTCAACCGGCTCAAGAAAGATGACAGCTCCAGCAGCTGAGTCAACCTGAC 3838
Db 565 oSerProProThrPheSerAlaPro-----ProProArgGlnIleHisSerPro----- 581
QY 3839 CTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3898
Db 582 -----ProProHisArgGln--ProArgProProT 592
QY 3899 CTCCACTTGTGAGAGCGCTCTGTATGCAACCTGTGAGCGAGGAGACAGAAAGC 3958
Db 592 hrProThrTyrGlyGlnPro--ProSerProProThrTyrTyrSerPro----- 607
QY 3959 TGTGAGCTCTGCGCCTTCTCTGAGCGGCTGGGCGTGTGATGCGATCAGCCCTCTCTCC 4018
Db 608 -----Pro 608
QY 4019 TCCCACTCTCCAAAGGTCTA 4039
Db 609 SerProProProTyrGlyLeu 615
```

RESULT 14

B35098
MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - human
C;Species: Homo sapiens (man)
C;Date: 10-Aug-1990 #sequence_revision 06-Nov-1992 #text_change 24-Aug-2001
C;Accession: B35098
R;Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.
Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990
A;Title: A gene pair from the human major histocompatibility complex encodes large protein
A;Reference number: A35098; MUID:90192810; PMID:2156268
A;Accession: B35098
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2142 <BAN>
A;Cross-references: GB:M33509; NID:gl79338; PIDN:AAA35585.1; PID:gl79339; GB:M31293
A;Note: the authors translated the codon AGT for residue 97 as Gly
C;Superfamily: collagen alpha 1(IV) chain

Alignment Scores:
Pred. No.: 1.28e-11 Length: 2142
Score: 313.50 Matches: 477
Percent Similarity: 29.88% Conservative: 178
Best Local Similarity: 21.76% Mismatches: 670
Query Match: 3.32% Indels: 869
DB: 2 Gaps: 119

US-09-931-704-3 (1-5087) x B35098 (1-2142)

```
QY 472 TGCCACATCTCTCTTACAGCACCCCTGGCCTCTGGACTCTGGTATCTCTGGAT 531
Db 39 CysProSerPro-----TrpPro 44
QY 532 GTCCAAACTCTGACG---TGCAATCAGCAACAAGCCGACTCGTCAAAATGACCTCTCT 588
Db 45 AlaGluSerArgGluSerCysHisCysProAla-----TyrArgPro 58
QY 589 CCCTTCTCTGCTCCCAACCTTGCAGGCTGATGGAAAGGCTCATATGAGTCCAACTTTTCC 648
Db 59 ProAlaAsnLeuProSerLeuLysAlaGluAsnLysGly-----Asn 72
QY 649 CCACCTACACAC-----AAGAACGGGGTGAACCTCCACACTGCCACCTGCTCC 696
Db 73 AspProAsnValSerLeuValProLysAspGlyThrGlyTrp----- 86
```

```
QY 697 TGAGAGTGAGCACTAATCTCTTCAATCTAAACCCCACTTCCACACTCCGACCTCAGGA 756
D 757 ATCAGATCTAGCACTAATCTCTTCAATCTAAACCCCACTTCCACACTCCGACCTCAGGA 756
D 87 -----AlaSerLysGlnGluInSerAsp----- 94
QY 757 ATCAGATCTAGCACTAATCTCTTCAATCTAAACCCCACTTCCACACTCCGACCTCAGGA 816
D 95 -----ProLysSerSerAspAlaSerThrAlaGlnProGluSer 108
QY 817 AACCTATACCTTGTCTTCCATGAGTGGTCTGAGTCTTCTTGGCGCGCTCTCTCTGCTT 876
D 109 GlnProLeuPro-----AlaSerGln--ThrProAla 119
QY 877 CTTCCCTTAGAGCTGACTGTGCTCAGCTGCCAGCTCTGACATGTGCTGCCACCT 936
D 119 erAsn-----GlnProLysArgProProAlaAla----- 130
QY 937 CTGACTCCCTCAGCTGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 989
D 130 luAsnThrProLeuValProSerGlyValLysSerTrpAlaGlnAlaSerValThrHisG 150
QY 989 ----- 989
D 150 lyAlaHisGlyAspGlyGlyArgAlaSerSerLeuLeuSerArgPheSerArgGluGluP 170
QY 990 -----CAACTGGAAACACAGGCGAGCTCGACC 1014
D 170 heProThrLeuGlnAlaAlaGlyAspGlnAspLysAlaAlaLysGluArgGluSerAlaG 190
QY 1015 TGAGTCCCTAGGCTGGCGCTCCCTCCATGATGATGATGATGATGATGATGATGATGATGAT 1074
D 190 luGlnSerSerGlyProGlyProSerLeuArgProGlnAsnSerThrTrpArgAspG 210
QY 1075 ACAGTGGCACACATGCCAAGTCTCTCAGCT--GACACACAGATCCATTCTCAAGTA 1131
D 210 lyGlyGly-ArgGlyProAspGluLeuGluGlyProAspSerLysLeuHis--HisGly 228
QY 1132 TCTACTGATAGACACTCATCGTGCCCAAGTCTCATCTCAACATACATACATGCTCTCT 1191
D 229 HisAspProArgGlyGlyLeuGlnProSerGlyProGln----- 242
QY 1192 TTTCTCTCCCTCTGCGAGAGTGTTCCTCTCCATCCCTCTGCTGCTGCTGCTGCTGCTGCT 1251
D 243 PheProProTyArgGlyMetMetProPheMetProPheMetProTyArgGlyProPhe 261
QY 1252 GTCCACCTCCACCCACCCACCCAGAGTGGGGGACAGACACTGAGGGGCTGCCAGC 1311
D 262 ProPro-ProTyGlyProGlnGlyProTyArgTyArgTyArgTyArgTyArgTyArgTy 281
QY 1312 TGCTTCCCGGTGTG-----GGCCCG----- 1331
D 281 rgPheProArgValAlaGlyProArgGlySerGlyProProMetArgLeuValGluProv 301
QY 1332 --GGCCGC----- 1337
D 301 aGlyArgProSerIleLeuLysGluAspAsnLeuLysGluPheAspGlnLeuAspGlnG 321
QY 1338 -----GCTCATGCT----- 1346
D 321 luAsnAspAspGlyTrpAlaGlyAlaHisGluGluValAspTyThrGluLysLeuLysP 341
QY 1347 -----TCTGCTCCATCTGCCCCACAGGAGTCTGTTGGGGATGTTAGC 1389
D 341 heSerAspGluGluAspGlyArgAspSerAspGluGluGlyAlaGluGlyHis-----A 359
QY 1390 GTGCCCTGTGACGGTGTCTGGACCTCCCTGAGTCCAGCTCTCAATCGCACAGGGA 1449
D 359 rgAspSerGlnSerAlaSerGlyGluGluArgProProGlu-----AlaAspGlyL 376
QY 1450 CCGAGGCTGGCCCTC----- 1470
D 376 yLysGlyAsnSerProAsnSerGluProProThrProLysThrAlaTrpAlaGluThrS 396
```

```
QY 1471 CCAGAAAAACCTATGACCTACCCGCTACCTGAGGACCACTCCGACGCTT----- 1521
D 396 erArgProGluThrGluProGly-ProProAlaProLysProProLeuProProGly 415
QY 1522 -----GGCTGGGACCTATGTAGTATCCAGCG-----TAGGAATCTGGG 1560
D 416 AspTyProAspArgGlyGlyProProCysLysProProAlaProGluAspGluAspGlu 435
QY 1561 AGTTGGGAGGAGTGGAG-----AGTTGGGGAAGACAGTCTCACTCCCTGGAGGTTCT 1614
D 436 AlaTrpArgGlnArgLysGlnSerSerSerGluLeuSerLeuAlaValGluArgAla 455
QY 1615 GGTAAATGATGGGTGAGAG----- 1635
D 456 ArgArgArgGluGluGluArgMetGlnGluArgAlaAlaCysAla 475
QY 1636 -----GGGCTCTTTGGCTCCACACAG----- 1656
D 476 GluLysLeuLysArgLeuAspGluLysPheGlyAlaProAspLysArgLeuLysAlaGlu 495
QY 1657 -----TCCCTCTGTCTCTATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1710
D 496 ProAlaAlaProProAlaAlaProSerThrProAlaPro-----ProProAlaVal 512
QY 1711 CCC-----ATCCCTGGCCCGCAGGACTAGGCATGTGGCGAGGCTCGCACCCCTTGGC 1764
D 513 ProLysGluLeuProAlaProAla-----ProProAlaSerAla 527
QY 1765 CCATTGCCCTAGCTGCTGCGAGCCGAGCCGCGCTCCCTTGGGGGCGGGGAGTCT 1824
D 528 ProThrProGluThrGluProGluProGluProAlaGlnAlaPro-----ProAlaGlnSer 545
QY 1825 CCTCTGTTTACACCTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1878
D 546 -----ThrProThrProGlyValAlaAlaAlaProThrLeuValSerGlyGly 562
QY 1879 CAGAGGGCCCCACCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1911
D 563 SerThrSerSerThrSerSerSerGlySerPheGluAlaSerProValGluProGlnLeuPro 582
QY 1912 GCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1971
D 583 SerLysGluGlyProGluPro-----ProGluGluValProPro-----Pro 596
QY 1972 TCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2031
D 597 ThrThrProProValProLysValGluProLysGly--AspGlyIleGlyProThrArg 615
QY 2032 GAACCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2091
D 616 GlnProProSer-----GlnGlyLeu-----GlyTyProLysTyGln 628
QY 2092 TGGGGTGTCTCAGAGTGTTCAGAGAGCCAG-----GAGAGAGG 2133
D 629 LysSerLeuProProArgPheGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 648
QY 2134 AAGGAGGTTGGAGGAGCCGAGGACCATGGGGAACCCGCCCCCTCTT----- 2181
D 649 GlnHisGlnTrpGlnHis-----GlnGlnGlySerAlaProProThrProValProPro 667
QY 2182 -----CCGCTGCTCTCTTCCACATCCAGACCTCTCTGAGGAGGAGGAGGAGGAGG 2229
D 668 SerProGlnProValThrLeuGlyAlaValProAlaProGlnAlaProProProPro 687
QY 2230 GAAAG-----GGAAGAGTGTGGGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2274
D 688 ProLysAlaLeuTyProGlyAlaLeuGlyArgProProProProProProProProPro 707
QY 2275 ACACGAGGAAATGTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328
D 708 AspProArgTrpMetMet-----IleProProTyVal 718
QY 2329 GGCCCTTGCCTATCCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2388
```

Db	719	AspProArgLeuLeuGlnGlyArgPro--ProLeuGluPheTyrProProGlyValHis	737
Qy	2389	CCCTCATCCCTT-----TCCCTGGGCCCCAGCGCTCCCCC--	2424
Db	738	ProSerGlyLeuValProArgGluArgSerAspSerLeuGlyLeuSerSerGluProPhe	757
Qy	2425	-----GAG	2427
Db	758	AspArgHisAlaProAlaMetLeuArgGluArgGlyThrProProValAspProLysLeu	777
Qy	2428	GGTTGGMAAGGCTCGCCCTCTTCCCTATACATGCTGCTTCATAGACCTTCCTCTG	2487
Db	778	AlaTrp-ValGlyAspValPheThrAlaThrPro--AlaGluProArgProLeuThrSer	796
Qy	2488	TCCTACTCATGAGACTGCC-----TCCATTTCCTCTTCTGCAACCCCTGC	2532
Db	796	rProLeuArgGlnAlaAlaAspGluAspAspLysGlyMetArgSerGluThrProProVa	816
Qy	2533	TCCTATCAGCTGAACCTCTCTTTCGGAGAGTGTAGTACGTACCCGCTCTCCACAGCCCT	2592
Db	816	lPro-ProProProTyrLeuAlaSer-----TyrProGlyPheProGluAsnG	833
Qy	2593	CAGCTGTGGG-----CTGGGTGTCTCAGCGCA	2622
Db	833	lyAlaProGlyProProIleSerArgPheProLeuGluGluProGlyProArgProLeuP	853
Qy	2623	AATGGGCTCGTTCCAATGGGCCACTCTCATCTCTCTTGTTCCTTGTGCAAGAAAC	2682
Db	853	roTrpProProGlySerAspGluValAlaLysIleGlnThrProProLysLysGluP	873
Qy	2683	CTTTGCTTCACTCCACTGCC-----CTCTCTAGTT	2712
Db	873	roProLysGluGluThrAlaGlnLeuThrGlyProGluAlaGlyArgLysLeuProAlaS	893
Qy	2713	CCCCA-----	2717
Db	893	erArgSerGlyAlaGlyProProProArgGluSerArgThrGluThrArgTrpG	913
Qy	2718	-----CCCTTTTCTCTCGCTTCCCTGCC	2745
Db	913	lyProArgProGlySerSerArgGlyIleProProGluGluProGlyAlaProProA	933
Qy	2746	AAATTTCT-----CCAAGAGTGGTCTACACCTCTGCTCCAC	2784
Db	933	rgArgAlaGlyProIleLysProProProThrLysValGluGluLeuProProL	953
Qy	2785	TTCCTCTCCAC-----CCACTCATCTTTAACCCCTGCAATCTGGTTCACAGCCCC	2837
Db	953	ysProLeuGluGlnGlyAspGluThrProLysProProLys-----ProAspProL	970
Qy	2838	-----AGCAATGGTCTCTCCAGGT-----CGTCAGCAGCTCTCT	2873
Db	970	euLysIleThrLysGlyLysLeuGlyGlyProLysGluThrProProAsnGlyAsnLeuS	990
Qy	2874	-----TGCC	2877
Db	990	erProAlaProArgLeuArgAspTyr-SerTyrGluArgValGlyProThrSerCysA	1010
Qy	2878	AAGCCCCA-----	2885
Db	1010	rgGlyArgGlyArgGlyGlyTyrPheAlaArgGlyArgGlyPheArgGlyThrTyrGlyG	1030
Qy	2886	-----CAGTGTTTCA	2897
Db	1030	lyArgGlyArgGlyGlyGlnAlaAsnSerAlaValThrGluSerPheGluGluMetMetG	1050
Qy	2898	-----GGCTCATTTCTGCTGCTGCTTTTGACGACACACTGCTGCTGAGCGTGC	2946
Db	1050	lyValGluValGlyGlnGlyAspGlnThrThrLeuLeuLeuProGlu-----AlaAlaM	1068
Qy	2947	TGCCTTCT-----	2954

[illegible]

QY 3649 CCCACTGCCCGCCGCTGCTGGGACTGAACCCACTTTGGACTCTCTGSCCTTGCCCA--- 3705
Db 1407 ArgThrGlyProGlyArgGly-----AspLysArgSerProSerProLys 1422
QY 3706 ---CAGTGAATCTCTCAGAAAGAT-----GGAGCACTTCTGGCTGCTGAAGGAGCT 3753
Db 1423 AsnArgSerArgProGluGluArgProGlyLeuProLeuProProProPro 1442
QY 3754 GCAGACTGCTGTG---GGCTCGGCCAA-----GGACTTCAA 3789
Db 1443 SerSerSerAlaValPheArgLeuAspGlnValIleHisSerAsnProAlaGlyIleGln 1462
QY 3790 CCGGCTCAAGAAAGATGACGCTCCAGCAGCTCGACTCCCTGACCTGGGGGCTCA 3849
Db 1463 GlnAlaLeuAlaGln---LeuSerSerArgGlnGlySerValThrAlaProGlyGlyHis 1481
QY 3850 TG-----GCTTCT 3857
Db 1482 ProArgHisLysProGlyProProGlnAlaProGlnGlyProSerProArgProProThr 1501
QY 3858 GACTTCTGACTTCTCTCTCTGCTCCCTCCCTTCAACCCCTGCTCCACTTTGTGAGACC 3917
Db 1502 ArgTyrGluProGlnArgValAsnSerGlyLeuSerSerAspProHisPheGluGluPro 1521
QY 3918 AGCCCTGTATGCCAACACCTGTTGAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTT 3977
Db 1522 GlyProMetValArgGlyValGlyGlyThrProArgAspSerAlaGlyValSerProPhe 1541
QY 3978 CTTGGACCGCTGGCGT-----GTGATCGATCAGCC 4010
Db 1542 Pro---ProLysArgArgGluArgProProArgLysProGluLeuLeuGlnGluSer 1560
QY 4011 CTGCTCTCTCCACCTCCCAAGGCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTG 4070
Db 1561 LeuProProHisSerSerGlyPheLeuGlySerLysProGluGlyProGlyProGln 1580
QY 4071 TCCTGTCTCTTCTACAGGAAGTCATGCTCGAGGGAGTGTGAAGTGGTTTCAGGTTG--- 4127
Db 1581 AlaGluSerArgAspThrGlyThrGluAlaLeuThrProHisIleTyrAsnArgLeuHis 1600
QY 4128 -----GTGCAGAGCGCTCATGCCCTCTGCTCTTCTGCTACCA----- 4166
Db 1601 ThrAlaThrSerArgLysSerTyrArgProThrSerMetGluProTyrMetGluProLeu 1620
QY 4166 ----- 4166
Db 1621 SerProPheGluAspValAlaGlyThrGluMetSerGlnSerAspSerGlyValAspLeu 1640
QY 4167 -----CTTGGCCAGTGC----- 4178
Db 1641 SerGlyAspSerGlnValSerSerGlyProCysSerGlnArgSerSerProAspGlyGly 1660
QY 4179 -----CCACCAGCCCTCAGTGGCAGCATCTGG-----AGG 4210
Db 1661 LeuLysGlyAlaAlaGluGlyProProLysArgProGlyGlySerProLeuAsnAla 1680
QY 4211 GCAGGGGTTGAGGGGCA-----CCACCACACATCCCTTTC 4246
Db 1681 ValProCysGluGlyProProGlySerGluProProArgArgProProProAlaProHis 1700
QY 4247 TGGGGTGAAGCCCTTGGCTGGCCCACTCTCTTGGATGGGTGTCTCCCTATATCCCA 4306
Db 1701 AspGlyAspArgLysGluLeuProArgGluGlnProLeuProProGlyProIleGlyThr 1720
QY 4307 AATCACTCTATACATCAATTCAGGAACAAACATGTTGGCAATTCACACAAAAGAGA 4366
Db 1721 GluArgSerGlnArgThrAspArgGlyThrGluProGlyProIleArgProSerHisArg 1740
QY 4367 TGAGATTAACAGTCAGGGTGGG-----GTCTGCATT 4399
Db 1741 ProGlyProProValGlnPheGlyThrSerAspLysAspSerAspLeuArgLeuValVal 1760

QY 4400 GGAGGTGCCCTATAAACACAGAGAGAAATACTGAAAGCACAGGGGCGGACAGACCAG 4459
Db 1761 GlyAspSerLeuLysAlaGluLysGluLeuThrAlaSerValThrGluAlaIleProVal 1780
QY 4460 ACCAGA-----CCGAGGAGTCTCCAAAGCACAGAGTGGCAACAAA--- 4501
Db 1781 SerArgAspTyrGluLeuLeuProSerAlaAlaAlaSerAlaGluProGlnSerLysAsn 1800
QY 4502 -----CCGAGCTGAGCATC-----AGGACCTTGCCTCGA 4531
Db 1801 LeuAspSerGlyHisCysValProGluProSerSerSerGlyGlnArgLeuTyrProGlu 1820
QY 4532 ATTGTCTTCCAGTATTACGGTCCCTTCTCTGCCCCCTTCCAGGCTATCTGTGGTT 4591
Db 1821 ValPheTyrGlySerAlaGlyProSerSer-SerGlnIleSerGlyGlySerHisGlyLe 1840
QY 4592 G-----CCAGGCTGG-----GGAGGGCAACCATATGCCACAC---CACAG 4627
Db 1840 uSerIleThrSerLysGlnTyrArgLeuArgProGlyThrProSerLeuHisProTyrAr 1860
QY 4628 GATTTCTCTGAAAGTTTACAATCAGTAGCATTTTGGGTGTAGGGTGGCAGCTCCCCAAG 4687
Db 1860 G-----SerGlnProLe 1864
QY 4688 GCCCTGCCCC-----CCAGCCCCCACCCTCATGACTCTAAGTGTGTGTTATTAATATT 4741
Db 1864 uTyrLeuProProGlyProAlaProProSer-----AlaLeuLeuSerG 1879
QY 4742 TATTTATTTGGAGATGTTATTTATTTAGATATATTTATTCAGAAATTTCTATTCTGTAT 4801
Db 1879 yValAlaLeuLysGlyGlnPhe-LeuAsp----- 1888
QY 4802 TAACAAATAAATGCTTCCGCCAGAACTTAGTCTCTTTGCCCGCAGC-----CTCACC 4852
Db 1889 --PheSerThrMetGlnAlaThrGluLeuGly-LysLeuProAlaGlyGlyValLeuTyr 1907
QY 4853 CTCTCTGTGTCTATCAGACTCTTCCACCCCTGGCTGCCACTCCCTGTGTGCT----- 4907
Db 1908 ProProProSer-PheLeuTyrSerProAlaPheCysProSerPro---LeuProAspTh 1926
QY 4908 ---CTGTGTGAGTGCACAGAGCTCTGGGA-----AGAGCCCTCT 4945
Db 1926 rSerLeuLeuGlnValArgGlnAspLeuProSerProSerAspPheTyrSerThrProLe 1946
QY 4946 TCCTCCC-----CGCACTGGGGCGATG-----GGCGCACCTCAGACTTACCCACTGCT 4993
Db 1946 uGlnProGlyGlyGlnSerGlyPheLeuProSerGlyAlaProAlaGlnGlnMetLeuLe 1966
QY 4994 G-----CTGCCA-----CCACCAACCCCTTG 5014
Db 1966 uProMetValAspSerGlnLeuProValValAsnPheGlySerLeuProProAlaProPr 1986
QY 5015 ATCCCTCAGTCTCCACACAGCTTCTGTCCACCCAGGTTTCCCTCACCCACCTTTCG 5074
Db 1986 oProAlaProProProLeuSerLeuLeu-ProValGlyProAlaLeuGlnProProSerL 2006
QY 5075 TAAGTCTT 5082
Db 2006 euAlaVal 2008
RESULT 15
S50832
atrophin-1 - human
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C;Accession: S50832
R;Nagatuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.;
Nature Genet. 8, 177-181, 1994
A;Title: Structure and expression of the gene responsible for the triplet repeat disc
A;Reference number: S50832; MUID:95144175; PMID:7842016
A;Accession: S50832
A;Status: preliminary
A;Molecule type: mRNA

[illegible]


```

Qy 5008 CCCCTTGATCCCTCAGTCCTCCACACAGCTTCTGTC-----CACCCCGAGTTTCCCTC 5061
    |||||:::|
Db 1095 nProLeuLeuProHisProLeuHisGluAsnGluValLeuArgHisGlnLeuPheAlaAl 1115
    |||||:::|
Qy 5062 ACCCCAC 5068
    :|||:::
Db 1115 aProTyr 1117
    :|||:::

```

Search completed: January 27, 2003, 16:27:02
Job time : 367.265 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:44:12 ; Search time 66.0255 Seconds
(without alignments)
6391.168 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aacctgcgagtgccctgcg.....ccttgtaagtctctctca 5087

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cg2_1/USPTO.spool/US09931704/runat_27012003_154125_3623/app_query.fasta_1.7189
-DB=SwissProt 40 -OFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09931704 @CGN 1.1 141 @runat_27012003_154125_3623 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	370.5	3.9	2161	1 SHK1_HUMAN	Q9Y566 homo sapien
2	351	3.7	2167	1 SHK1_RAT	Q9W48 rattus norv
C 3	327.5	3.5	1466	1 CA13_HUMAN	P02461 homo sapien
C 4	325	3.5	1464	1 CA13_MOUSE	P08121 mus musculu
C 5	317.5	3.4	1262	1 CA13_CHICK	P12105 gallus gall
6	316	3.4	1685	1 CA54_HUMAN	P29400 homo sapien
7	314.5	3.3	620	1 EXTN_TOBAC	P13983 nicotiana t
8	313.5	3.3	2142	1 BAT2_HUMAN	P48634 homo sapien
9	310	3.3	1690	1 CR44_HUMAN	P53420 homo sapien
C 10	306.5	3.3	1362	1 CR21_CHICK	P02467 gallus gall
C 11	306.5	3.3	1453	1 CA11_MOUSE	P11087 mus musculu
C 12	306.5	3.2	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
C 13	305	3.3	1418	1 CA12_HUMAN	P02458 homo sapien
C 14	304.5	3.3	1049	1 CA13_BOVIN	P04258 bos taurus
C 15	304.5	3.2	1183	1 DRPL_RAT	P54258 rattus norv
C 16	304.5	3.3	1355	1 CR21_PANCA	O42350 rana catesb
C 17	303	3.2	2944	1 CA17_HUMAN	Q02388 homo sapien
C 18	302.5	3.2	1255	1 MUC1_HUMAN	P15941 h mucin 1 p

Q9nzm4	homo sapien	1509	3.2	302	1 GSRI_HUMAN
P39060	homo sapien	1516	3.2	297.5	1 CA1H_HUMAN
Q01955	homo sapien	1670	3.1	296.5	1 CA34_HUMAN
P53420	homo sapien	1690	3.2	296.5	1 CA44_HUMAN
P54259	homo sapien	1185	3.1	295.5	1 DRPL_HUMAN
Q9xsj7	canis fami	1460	3.2	295	1 CA11_CANFA
P08123	homo sapien	1366	3.1	293.5	1 CA21_HUMAN
P05997	homo sapien	1496	3.1	292	1 CA25_HUMAN
O93484	oncorhynch	1356	3.1	292	1 CA21_ONCMY
P28481	mus musculu	1459	3.1	291.5	1 CA12_MOUSE
P02452	homo sapien	1464	3.1	291.5	1 CA11_HUMAN
Q9wu42	mus musculu	2472	3.1	291	1 NCR2_MOUSE
P02457	gallus gall	1453	3.1	290	1 CA11_CHICK
P02466	rattus norv	1372	3.1	289.5	1 CA21_RAT
P20908	homo sapien	1838	3.0	284.5	1 CA15_HUMAN
Q01149	mus musculu	1372	3.0	284	1 CA21_MOUSE
P33485	pseudorabie	1733	3.0	283	1 VNUA_PRVKA
O46392	canis fami	1366	3.0	281.5	1 CA21_CANFA
Q81245	mus musculu	1804	3.0	281.5	1 CA1B_MOUSE
P27393	ascaris suu	1763	3.0	279.5	1 CA24_ASCSU
Q9fpq6	chlamydomon	555	3.0	278.5	1 GP1_CHLRE
Q07092	homo sapien	1603	2.9	277.5	1 CA1F_HUMAN
Q07092	homo sapien	1806	3.0	277.5	1 CA1B_HUMAN
P32107	homo sapien	817	2.9	275	1 VRP1_YEAST
P37370	saccharomyc	1603	2.9	275	1 CA1F_HUMAN
Q07092	homo sapien	1670	2.9	274	1 CA34_HUMAN
Q01955	homo sapien	1496	2.9	273.5	1 CA25_HUMAN
P05997	homo sapien				

ALIGNMENTS

RESULT 1

SHK1_HUMAN	SHK1_HUMAN	STANDARD;	PRT;	2161 AA.
AC	Q9Y566; Q9NYW9;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DE	SH3 and multiple ankyrin repeat domains protein 1 (Shank1)			
DE	(SSTRIP).			
DE	(SSTRIP).			
GN	SHANK1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND INTERACTION WITH SSTR2.			
RC	TISSUE=Fetal brain, Hippocampus, and Thalamus;			
RX	MEDLINE=20020275; PubMed=10551867;			
RA	Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;			
RT	"Somatostatin receptor interacting protein defines a novel family of			
RT	multidomain proteins present in human and rodent brain.";			
RL	J. Biol. Chem. 274:32997-33001(1999).			
RN	[2]			
RP	REVIEW.			
RA	PubMed=10806096;			
RA	Sheng M., Kim E.;			
RT	"The Shank family of scaffold proteins.";			
RL	J. Cell Sci. 113:1851-1856(2000).			
CC	-!- FUNCTION: Seems to be an adapter protein in the postsynaptic			
CC	density (PSD) of excitatory synapses that interconnects receptors			
CC	of the postsynaptic membrane including NMDA-type and metabotropic			
CC	glutamate receptors via complexes with GKAP/PSD-95 and Homer,			
CC	respectively, and the actin-based cytoskeleton. May play a role in			
CC	the structural and functional organization of the dendritic spine			
CC	and synaptic junction.			
CC	-!- SUBUNIT: May homomultimerize via its SAM domain (By similarity).			
CC	Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with			
CC	SPTAN1, Homer-1 and DLGAP1/GKAP isoforms 1 and 2 (By similarity).			
CC	Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By			
CC	similarity).			
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of			


```

CC DR EMBL; AF102855; AAD04569.2; --
DR EMBL; AF131951; AAD29417.1; ALT_INIT.
DR EMBL; AF159045; AAD42975.1; --
DR EMBL; AF141904; AAF02498.1; ALT_INIT.
DR HSSP; P00519; IABL.
DR DR InterPro; IPR002110; ANK.
DR DR InterPro; IPR001478; PDZ.
DR DR InterPro; IPR001660; SAM.
DR DR InterPro; IPR001452; SH3.
DR DR Pfam; PF00023; ank; 7.
DR DR Pfam; PF00595; PDZ; 1.
DR DR Pfam; PF00536; SAM; 1.
DR DR Pfam; PF00018; SH3; 1.
DR DR ProDom; PD000066; SH3; 1.
DR DR SMART; SM00248; ANK; 3.
DR DR SMART; SM00228; PDZ; 1.
DR DR SMART; SM00454; SAM; 1.
DR DR SMART; SM00326; SH3; 1.
DR DR PROSITE; PS0088; ANK_REPEAT; 3.
DR DR PROSITE; PS0297; ANK_REPEAT; 1.
DR DR PROSITE; PS0002; SH3; 1.
DR DR PROSITE; PS0105; SAM_DOMAIN; 1.
DR DR ANK repeat; SH3 domain; Repeat; Alternative splicing.
KW ANK repeat; SH3 domain; Repeat;
FT REPEAT 195 210 ANK 1.
FT REPEAT 212 245 ANK 2.
FT REPEAT 246 278 ANK 3.
FT REPEAT 279 312 ANK 4.
FT REPEAT 313 345 ANK 5.
FT REPEAT 346 378 ANK 6.
FT REPEAT 379 395 ANK 7.
FT DOMAIN 554 613 SH3.
FT DOMAIN 663 757 PDZ.
FT DOMAIN 2104 2167 SAM.
FT DOMAIN 929 932 POLY-PRO.
FT DOMAIN 1010 1015 POLY-HIS.
FT DOMAIN 1022 1027 POLY-HIS.
FT DOMAIN 1194 1199 POLY-GLY.
FT DOMAIN 1850 1860 POLY-PRO.
FT VARSPLIC 1 613 MISSING (IN ISOFORM 2).
FT VARSPLIC 615 654 SOEGROESRDKAKRLFRHYTVGSYDSFAPSLIDGDSG
FT -> MALSAVGGPGGGALPQPPALSSSWPALGPRRRSVY
FT IY (IN ISOFORM 2).
FT VARSPLIC 646 654 MISSING (IN ISOFORM 3).
FT VARSPLIC 797 804 MISSING (IN ISOFORM 4).
FT VARSPLIC 1930 1943 LSEDSTLSKPS -> QYRIIVKSSDFGDF (IN
FT ISOFORM 5).
FT VARSPLIC 1944 2167 MISSING (IN ISOFORM 5).
FT CONFLICT 1141 1141 S -> T (IN REF. 1).
FT CONFLICT 1174 1174 S -> N (IN REF. 2).
FT CONFLICT 1246 1246 R -> K (IN REF. 1).
FT CONFLICT 1323 1323 A -> T (IN REF. 1).
FT CONFLICT 1331 1331 S -> D (IN REF. 1).
FT CONFLICT 1726 1726 S -> N (IN REF. 2).
SQ SEQUENCE 2167 AA; 226333 MW; 3F478B5A7B18BA86 CRC64;

Alignment Scores:
Pred. No.: 4,98e-09 Length: 2167
Score: 351.00 Matches: 334
Percent Similarity: 29.90% Conservative: 102
Best Local Similarity: 22.91% Mismatches: 483
Query Match: 3.72% Indels: 541
DB: 1 Gaps: 69

US-09-931-704-3 (1-5087) x SHK1_RAT (1-2167)
QY 1084 CACATGCCAAAGACTCTC-----TCAGCTGACACACAGATCCATTCACAA 1128
Db |||||
Db 858 HisArgProLysGlyPhePheAlaThrGluSerSerPheAspProHisHisArgSerGln 877
QY 1129 GTATCTACTAGACACTCATCGTGCCA-----|||||

```

```

Db 878 ProSerTyrAspArgProSerPheLeuProProGlyProGlyLeuMetLeuArgGlnLys 897
QY 1159 -----AGTCCTCATCTCTCAAAACATACACA 1182
Db |||||
Db 898 SerIleGlyAlaAlaGluAspArgProTyrLeuAlaProProAlaMetLysPheSer 917
QY 1183 TGCCTCTCTTCTCTCCCGTCTTCCAGAGTGTTCCTCTCTCCATCC-----C 1233
Db |||||
Db 918 ArgSerLeuSerValProGlySerGluAsp-IleProProProThrThrSerProP 937
QY 1234 CTCTGCTCTCCATCTGGTGTCTCCACCTCACCCACCCAGTCCAGTGGGGACAGAC 1293
Db |||||
Db 937 roGluProProTyr-SerThrProAlaProSerSerGlyGlyPro-LeuProAlaSerSerP 956
QY 1294 ACCTGAGGGGCTGCAGCTGCTTCCCGTGTGGGCGCGCGCTCATGCTTCTCGTC 1353
Db |||||
Db 956 rProArgGlyGlyPro---PheAsnProSerSerGlyGlyPro-LeuProAlaSerSerP 975
QY 1354 CATCC-----TGCCACACAGGGACTCTGGGGGATGTTAGCTGTGCT--- 1395
Db |||||
Db 975 roSerSerPheAspGlyProSerProProAspThrArgGlyGly-GlyArgGluLysSer 994
QY 1396 ---GTGCACGGTGTCTGGCACCTCCCTGC-----AGTGCCAGCTCTCAATCGCAC 1443
Db |||||
Db 995 LeuTyrHisSerAlaAlaLeuProProAlaHisHisHisProProHisHisHis 1014
QY 1444 AGGGGACCCAGGGCTGGCCCTCCATCCAGAAAACCTATGACCTCACCCG----- 1494
Db |||||
Db 1015 HisAlaProProProGlnProHisHisHisAlaHisProProHisProProGluMet 1034
QY 1495 -----CTACCTGGAGCACCACTCCGACGTTGG-CTGGGACCTATGTAGTA 1541
Db |||||
Db 1035 GluThrGlyGlySerProAspAspProProProArgLeuAlaLeuGlyProGln----- 1052
QY 1542 TCCAGCTAGGAATCTGGAGTGGGAGGAGTAGGAGTGGGGAAAGACAGCTCTAAC 1601
Db |||||
Db 1053 ProSerLeuArgGlyTrp----- 1058
QY 1602 CGTGGAGGGTCTGTAAATGATGGGGTGAGGAGGGCTCTTTGGCTCCACACAGTCC 1661
Db |||||
Db 1059 ArgGlyGly-----GlyProSerPro 1065
QY 1662 CTGTCTGTCTATCTCTGCTCCCTTCCCTTCTTAGTGGCCCCCCCCACTTCCCCATCCCTGG 1721
Db |||||
Db 1066 ThrSerGly-----AlaProSerProSerHisSer 1076
QY 1722 CCCAGGACTAGGATGTGGGCGCCCTCGCACCCGCT-----TGGCCCATGGCC 1772
Db |||||
Db 1077 SerSerGlyGlySerSerSerGlyProThrGlnAlaProAlaLeuArgTyrPheGlnLeuPro 1096
QY 1773 CCAGTGGTGGCAGCCAGCCGCGCTCCCTGGGGCGCGGAGTCTCTCTCTGT 1832
Db |||||
Db 1097 ProArgAlaAlaSerAlaAlaMetTyrValProAlaArgSerGlyArgGlyGly 1116
QY 1833 TACACCGTGTGTGTCTCTTCTCGCGGGGGGTGGTGGGGGACA----- 1880
Db |||||
Db 1117 ProLeuValLysGlnThrLysValGluGlyGluProGlnLysGlySerIleProSerAla 1136
QY 1881 GAGGGGCCCCACCTCCCATGCTGCTGCTCCAGTCCGCTCTCCCCACACCTGGGGCCC 1940
Db |||||
Db 1137 SerSerProThrSerProAlaLeuProArgSerGlu-----ProProProAlaGlyPro 1154
QY 1941 TGCTGCTCTGACCCAGGGGCTCCCTTCCGCTCTCCCTCTCCATCTCCAGTGGGGCTCC 2000
Db |||||
Db 1155 SerGluLysAsnSerIleProIleProThrIleIleLysAlaProSerThrSerSer 1174
QY 2001 TAGGGGGTCTATGGGGGAAGGGGACTGTAGGGA-----ACC 2036
Db |||||
Db 1175 -----SerGlyArgSerSerGlnGlySerSerThrGluAlaGluProProThr 1190
QY 2037 CAGCAGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGGTTATGCTGTAAAGATTGGGG 2096
Db |||||
Db 1191 GlnProAspGlyAlaGlyGlyGly----- 1199

```



```
Db 1855 roProLeuProGlyProLeuSerGlnProGlnAlaSerAlaLeuAlaThrValLysAla 1875
QY 3603 -----
Db 1875 erIlelleSerGluLeuSerSerLysLeuGlnGlnPheGlyGlySerThrAlaGlyG 1895
QY 3604 -----GGGCGCTGCTGGCGCAGCATGTCGGGGCTCATGCGC----- 3636
Db 1895 lyAlaLeuProTtpAlaArgGlyGlySerGlyGlySerThrAspSerHisGlyGlyA 1915
QY 3637 -----A 3637
Db 1915 laSerTyrIleProGluArgThrSerSerLeuGlnArgGlnArgLeuSerGluAspSerG 1935
QY 3638 GCCTGGCTTACCCAGTCCGCCCGCCCGCTGCTGGAGTGAACCCACTTGACTCTGGC 3697
Db 1935 lnThrSerLeuSerLysProSerSer-----SerIlePheGlnAsnTrpP 1951
QY 3698 CTTGCCACAGTACTTCTCCAGAGATGAGACGACTTCTGGCTGCTGAAGGAGCTGCAG 3757
Db 1951 roLysProPro-----LeuProPro-----LeuProThrGlySerGlyValSerS 1966
QY 3758 ACCTGGCTGTGGCGCTCGGCCAGGACTTCAACCGGCTCAAGAAAGATCCAGCTCCA 3817
Db 1966 erSerThrAlaAlaAlaProGlyAlaThrSerProSer-----AlaSerSerAlaSerA 1984
QY 3818 GCAGCTGCAGTCACTCCCTGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3877
Db 1984 laSerThrArgHis-----LeuGlnGlyValGluPheGluMetArgProProLeuLeuArgA 2003
QY 3878 TCCTCTCCC-----CTTCA-----AACCTCTCTCCCACTT 3907
Db 2003 rgAlaProSerProSerLeuLeuProAlaSerAspHisLysValSerProAlaProArg- 2022
QY 3908 TGTGAGGAGCAGCCCTGTATGCCACACA---CCTGTTGAGCCAGCAGACAGAGCTGTGAG 3964
Db 2023 ProSerSerLeuProLeuLeuProSerGlyProIleTyrProGlyLeuPheAspIleArg 2042
QY 3965 CCTCTGCGCCCTTCTGCGACCGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4024
Db 2043 SerSerProThrGlyGlyAlaGlyGlySerThrAspProPheAlaProValPheValPro 2062
QY 4025 CTTCCCAAGAGTTCAGGAGCTGGGAGGAGGTACAGTAGCCCTGCTGCTGCTGCTGCTGCT 4084
Db 2063 ProHisProGlyIle-----SerGlyGlyLeuGlyGlyAlaLeuSer----- 2076
QY 4085 TACAGGAAGTCATGCTCGAGGAGTGTGAAGTGTGAGTGTGAGTGTGAGGAGGCTCATG 4144
Db 2077 -----GlyAlaSerArgSerLeuSerProThrArgLeu----- 2087
QY 4145 GCCTCTGCTTCTTGCTTACCACTTGGCCAGTGGCCAGTGGCCAGTGGCCAGTGGCCAGT 4204
Db 2088 -----LeuSerLeuProProAspLysProPheGlyAlaLysProLeu----- 2101
QY 4205 TGGAGGGCAGGGGTTGAGGGGCCACACACATGCTTCTTGGGGTGAA----- 4255
Db 2102 -----GlyPheTrpThrLysPheAspVal 2109
QY 4256 GCCTTTTGGCTGCCCACTCTCTCTGATGGGTGTGCTCTCCCTTATCCCAATCACTCT 4315
Db 2110 AlaAspTrpLeu-----GluTrpLeuGlyLeuSerGluHisArgAlaGlnPheLeu 2126
QY 4316 ATACATCCATCAGGAACAACATGTTGGCAATT 4351
Db 2127 AspHisGluIleAspGlySerHisLeuProAlaLeu 2138
RESULT 3
CA13 HUMAN
ID CA13_HUMAN STANDARD; PRT; 1466 AA.
AC P02461; Q15112;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin fibroblast;
RX MEDLINE=89350838; PubMed=2764886;
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT chain of human type III procollagen. Differences in protein structure
RT from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516 (1989).
RN [2]
RP SEQUENCE OF 149-1225 FROM N.A.
RX MEDLINE=89386015; PubMed=2780304;
RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1
RT (III) collagen.";
RL Nucleic Acids Res. 17:6742-6742 (1989).
RN [3]
RP SEQUENCE OF 168-398.
RX MEDLINE=77134724; PubMed=5573335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanogen
RT bromide peptides from the amino-terminal segment of type III collagen
RT of human liver.";
RL Biochemistry 16:1158-1164 (1977).
RN [4]
RP REVISIONS.
RA Seyer J.M.;
RL Submitted (DEC-1977) to the PIR data bank.
RN [5]
RP SEQUENCE OF 399-727.
RX MEDLINE=79000343; PubMed=687591;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBr peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411 (1978).
RN [6]
RP SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589 (1980).
RN [7]
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Daigleish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337 (1988).
RN [8]
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Daigleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833 (1988).
RN [9]
RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Leidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394 (1984).
RN [10]
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
```

RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha
RT 1(III)-C99 from type III collagen of human liver.";
RL Biochemistry 20:2621-2627(1981).
RN [11]
RP SEQUENCE OF 1176-1466 FROM N.A.
RX MEDLINE=85157600; PubMed=2579949;
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT (III) collagen. Partial characterization of the 3' end region of the
RT gene.";
RL J. Biol. Chem. 260:4357-4363(1985).
RN [12]
RP SEQUENCE OF 1161-1200 FROM N.A.
RX MEDLINE=86187804; PubMed=3754462;
RA Miskulin M., Dalgleish R., Klueve-Beckerman B., Rennard S.I.,
RA Tolstoshev P., Brantly M., Crystal R.G.;
RT "Human type III collagen gene expression is coordinately modulated
RT with the type I collagen genes during fibroblast growth.";
RL Biochemistry 25:1408-1413(1986).
RN [13]
RP SEQUENCE OF 1-170 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88303360; PubMed=3405773;
RA Toman D., Ricca G., de Crombrughe B.;
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT of human prepro alpha 1(III) collagen.";
RL Nucleic Acids Res. 16:7201-7201(1988).
RN [14]
RP SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=89378752; PubMed=2777083;
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT "Cloning and analysis of the 5' portion of the human type-III
RT procollagen gene (COL3A1).";
RL Gene 78:255-265(1989).
RN [15]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [16]
RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX MEDLINE=93293988; PubMed=8514866;
RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
RA Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.T.,
RA Majamaa K., Smullen S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT in the triple-helical domain of type III procollagen are an
RT infrequent cause of aortic aneurysms.";
RL J. Clin. Invest. 91:2539-2545(1993).
RN [17]
RP VARIANT THR-698.
RX MEDLINE=91045136; PubMed=2235526;
RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA Wu Y., Ganguly A., Prockop D.J.;
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL Nucleic Acids Res. 18:6180-6180(1990).
RN [18]
RP VARIANT AORTIC ANEURYSM ARG-786.
RX MEDLINE=91056145; PubMed=2243125;
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT "A mutation in the gene for type III procollagen (COL3A1) in a family
RT with aortic aneurysms.";
RL J. Clin. Invest. 86:1465-1473(1990).
RN [19]
RP VARIANT EDS-IV ARG-828
RX MEDLINE=94016385; PubMed=8411057;
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;

RT "The substitution of glycine 661 by arginine in type III collagen
RT produces mutant molecules with different thermal stabilities and
RT causes Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 30:690-693(1993).
RN [20]
RP VARIANT EDS-IV SER-957.
RX MEDLINE=89109135; PubMed=2492273;
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
RT "A single base mutation that substitutes serine for glycine 790 of
RT the alpha 1 (III) chain of type III procollagen exposes an arginine
RT and causes Ehlers-Danlos syndrome IV.";
RL J. Biol. Chem. 264:1349-1352(1989).
RN [21]
RP VARIANT EDS-IV VAL-960.
RX MEDLINE=95268429; PubMed=7749417;
RA Tromp G., de Paeppe A., Nuytink L., Madhathari S.L., Kuivaniemi H.;
RT "Substitution of valine for glycine 793 in type III procollagen in
RT Ehlers-Danlos syndrome type IV.";
RL Hum. Mutat. 5:179-181(1995).
RN [22]
RP VARIANT EDS-IV GLU-1014.
RX MEDLINE=92316511; PubMed=1352273;
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA Pope F.M.;
RT "A single base mutation in the gene for type III collagen (COL3A1)
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT syndrome type IV. An unaffected family member is mosaic for the
RT mutation.";
RL Hum. Genet. 89:414-418(1992).
RN [23]
RP VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT "Single base mutation in the type III procollagen gene that converts
RT the codon for glycine 883 to aspartate in a mild variant of
RT Ehlers-Danlos syndrome IV.";
RL J. Biol. Chem. 264:19313-19317(1989).
RN [24]
RP VARIANT EDS-IV VAL-1077.
RX MEDLINE=91374480; PubMed=1895316;
RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
RA Pope F.M.;
RT "Characterisation of a glycine to valine substitution at amino acid
RT position 910 of the triple helical region of type III collagen in a
RT patient with Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 28:458-463(1991).
RN [25]
RP VARIANT EDS-IV GLU-1173.
RX MEDLINE=93022543; PubMed=1357232;
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
RT Alignment Scores:
Pred. No.: 5.65e-08 Length: 1466
Score: 327.50 Matches: 326
Percent Similarity: 29.96% Conservative: 80
Best Local Similarity: 24.06% Mismatches: 501
Query Match: 3.51% Indels: 450
DB: 1 Gaps: 71
US-09-931-704-3 (1-5087) x CAL3_HUMAN (1-1466)
Qy 5059 GGGAAACCTGGGGTGGACAGAAAGCTGTGTGGAGGACTGAGGGATCAAGGGTGTGGTGT 5000
Db 216 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysaspGly 231
Qy 4999 GGCAGCAGCAGTGGTAAAGTCTGAGGTGGCCCTCGCCAGTGGGGGAGAGAGGG 4940
Db 232 GluSer-----GlyArgProGlyArgPro-----GlyGluargGly 243
Qy 4939 ---CCTCTTCCAGAGCTCTGTGCAGCTCCACGAGGCAAGCAGGAGTGTG----- 4892
Db 244 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 259
Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;

"Identification of the promoter and first exon of the mouse alpha 1 (III) collagen gene.";
 J. Biol. Chem. 260:3773-3777(1985).
 [4]
 SEQUENCE OF 810-1464 FROM N.A.
 STRAIN=C57BL/6J; TISSUE=Embryonic head;
 MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukuishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 Radocha K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 Nature 409:685-690(2001).
 [5]
 SEQUENCE OF 1442-1464 FROM N.A.
 STRAIN=C57BL/6;
 MEDLINE=91274355; PubMed=2054384;
 RA Metsieranta M., Toman D., de Crombrughe B., Vuorio E.;
 RA "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs.";
 RT Biochim. Biophys. Acta 1089:241-243(1991).
 CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 CC ALONG WITH TYPE I COLLAGEN.
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC O-LINKED GLYCANS CONSISTS OF GLC-6AL DISACCHARIDE (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 1 VFWC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X52046; CAA36279.1; -;
 DR EMBL; M18933; AAA37338.1; -;
 DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
 DR EMBL; AK019448; BAB31724.1; -;
 DR EMBL; X57983; CAA41048.1; -;
 DR PIR; A22287; A22287.
 DR PIR; A27353; A27353.
 DR PIR; S16373; S16373.
 DR MGI; 88453; Col3a1.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01391; Collagen_18.
 DR Pfam; PF01410; COLFI_1.
 DR ProDom; PD000007; Collagen; 1.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VFWC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90 VFWC.
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION (C-TERMINAL).
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).
 FT CARBOHYD 262 262 O-LINKED (GAL. .) (BY SIMILARITY).
 FT MOD RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT MOD RES 976 976 HYDROXYLATION (BY SIMILARITY).
 FT MOD RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
 FT MOD RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;
 Alignment Scores:
 Pred. No.: 7.29e-08 Length: 1464
 Score: 325.00 Matches: 377
 Percent Similarity: 28.83% Conservative: 73
 Best Local Similarity: 24.15% Mismatches: 549
 Query Match: 3.48% Indels: 562
 DB: 1 Gaps: 83
 US-09-931-704-3 (1-5087) x CA13_MOUSE (1-1464)
 QY 4672 CCTACACCCCAAAATGCTACTGCTTAACTTTTCAGGAATCTGTGGTGGCTAT 4613
 DB 92 ProSerThrPro-----AlaProValLeuProAspGlyHis 103
 QY 4612 GGTTCCTCCCTCCAG-----CTGGCAACCCACAGATACCTCGGGAAGG 4568
 DB 104 Gly-----ProGlnGlyProLysGlyAspProGlyProGlyIleProGlyAsn 121
 QY 4567 GGGCAGAGAGAGGACCGCTAATACTGGAAGACATTCGAGGCAAGTCTGTGCTCAG 4508
 DB 122 GlyAsp-----ProGlyLeuProGlyGln 129
 QY 4507 CTCGGGTTTGTTCCTCCTCTGCTGCTGGAGACTCTCGGGTCTGCTGGTCTGT--- 4451
 DB 130 ProGlyLeuProGlyPro-----ProGlySerProGlyIleCysGlu 143
 QY 4450 CCTTCCTCCCTGTGCTTTCAGTATTTCTCTTCGTTTATAGGCACTCCCAATGAGAC 4391
 DB 144 SerCysProThr-----GlyGlyGlnAsnTyrSer 153
 QY 4390 CCCAACCTCGCACTGTAACTCTCTCTTTTGTGTAGAAATGCCACCATGTTGTTC 4331
 DB 154 ProGln-----PheAspSer----- 158
 QY 4330 CTGAATTGGATGTATAGAGTGTATTTGGGATAAGGAGCAACACCATCCAGGAGGTG 4271
 DB 159 -----TyrAspValLysSerGly-----Val 165
 QY 4270 GGGCAGCCAAAGGCTTCACCCCAAGAGGATGTGTGGTGGCCCTCAACCCCTGC 4211
 DB 166 GlyGlyMetGlyGlyTyrPro-----GlyProAlaGlyProPro 178
 QY 4210 CCTCCAGATGTGCCACCTGAGGGGCTGGTGGGCTAGTGGCAAGTGGTAGGCAAGCA 4151
 DB 179 GlyProProGlyProPro---GlySerSerGlyHis-----ProGlySerProGlySer 195
 QY 4150 GGAGGCCATGAGCGCTCTCGACCAACCACTTCACACTTCACACTCCCTCGAGCATGACTT 4091
 DB 196 ProGlyTyrGlnGly-----Pro 201
 QY 4090 CCTGTAGAACAGGACAGGAGGCGCTACTGTACTCTCTCCCTCCAGCTGGTAGACTTT 4031
 DB 202 ProGlyGluProGlyGln-AlaGlyProAlaGlyProGlyProGly---AlaLeu 220

QY 3973 GGCAGAGGCTCACAGCTCTCTG-----TCTCTGGCTCAACAGGTGTT-- 3931
Db 239 YAspArgGlyLeuProGlyLeuProGlyHisLysGlyHisProGlyMetProGlyMetPro 259
QY 3930 -GGCATACAGGCTGGCTCTCACAAGTGGAG-CAGGGTTGAAGGGGAGCGAGAGG 3873
Db 259 oGlyMetLysGlyAlaArgGlyPheAspGlyLysAspGlyAlaLysGlyAspSerGlyAl 279
QY 3872 AGAAGTCAAGTCAGAAGCCATGAGCCCGCAGGTGAGGTGACTGCAGCTGCTGAG 3813
Db 279 aProGlyProLysGlyGluAlaGlyGlnProGlyAlaAsnGlySer----- 294
QY 3812 GCTGCATCTTCTTCTGAGCGGTGAAGTCTCTGGCGGCGCCACAGCCAGGTCT-GC 3754
Db 295 -----ProGlyGlnProGlyProGly 301
QY 3753 AGCTCTTACAGCCAGAGTCTCATCTTCTGGAGAGTCACTGTGGCGAGGCCA 3694
Db 301 yGlyProThrGlyGluArgGlyArgProGlyAsnProGlyGlyProGlyAlaHisGlyLy 321
QY 3693 GGAGTCCAAGTGGTTCAGTCCAGCAGCGCTGGCGAGTGGTGGTGGTGGTGGTGGTGG 3634
Db 321 sAspGly-----AlaProGlyThrAlaGlyProLeuGlyProGlyProPr 337
QY 3633 ATGACGCCCAATGCTGCCAGCAGCGCTGGAGGCTGTGCAAGTGGTGGTGGTGGTGGTGG 3574
Db 337 o---GlyThrAlaGlyPheProGlySerProGlyPheLysGlyGluAlaGlyPro----- 354
QY 3573 CGCGCGAGCTCAGCAGTGGCAGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3514
Db 355 -----ProGlyProAlaGlyAlaSerGlyAsnProGlyGluAr 367
QY 3513 CTGTAGGCTCTAGTCTGGTTCAGCGCGCATTTGTTCATTGAGGCTTCGCGCACACTCC 3454
Db 367 gGlyGluProGlyProGlnGlyGlnAlaGly-----ProPr 379
QY 3453 AAGTCAACAGTGGCTGGCAGAGTCTCTGCCCGCGGAGGAGGTGAAGTCTGGC 3394
Db 379 oGlyProGlnGlyProProGlyArgAlaGlySerProGlyGlyLysGly--GluMetGly 398
QY 3393 -----TCGTTGAAGGGGCGCAGGTAGTTCAGCTGTGAAAGGAGAGGTATGGGG 3340
Db 399 ProSerGlyLeuProGlyGlyPro-----Gly 407
QY 3339 AAGGAGGAGGCGCAGCGCTGGCTCACCACAGGAGATACCTGCTCCCAAGTTCTA 3280
Db 408 ProProGlyGlyArgGlyLeuProGlyProPro----- 418
QY 3279 TTTTGTGTCCCTGACACTGGCCCTGTCTCCAGTAGGCTTCTTTGTTATGGCCTC 3220
Db 418 ----- 418
QY 3219 CCTGAGTTTTCATAAGGATATCATTTGTATGTCAGGTGACAGTTGAAGCAGTGGCGGT 3160
Db 419 -----GlyThrSerGlyAsnProGly-Al 426
QY 3159 GGTGAGCTCTCAATGAGAGGGGTAGAAAAGGAGGAGGAGAGAC---GCACAAGAGC 3103
Db 426 aLysGlyThrProGlyGluProGlyLys-AsnGlyAlaLysGlyAspProGlyProLysG 446
QY 3102 CTCAGAAGAGCTGCTGATGGCTGGTGGGCGAAACCTTGGGAGTAGGCCCGCTGTGGG 3043
Db 446 lyGluArgGlyGluAsnGlyThrProGlyAlaArg-----G 458
QY 3042 GGCAGGACAGAGAGGAGAGGCTCGAGGAGCTGGAGAG-----GTAG 2998
Db 458 lyProProGlyGluGluGlyLysArgGlyAlaAsnGlyGluProGlyGlnAsnGlyValP 478
QY 2997 AGGTGGCCGAGAGTGCAGACACAAGG-----AGAGGAGTT----- 2957
Db 478 roGlyThrProGlyGluArgGlySerProGlyPheArgGlyLeuProGlySerAsnGlyL 498

QY 2956 -----CGAAGAGCAGCAGCGCTCAGCAGTGTGGCTGCACAAAACAGACAGCAAGAGAAATG 2902
Db 498 euProGlyLysGlyProAlaGlyGluArgGlySerProGlyProProGlyProSerG 518
QY 2901 AGCTTCAAAAACACTGTTCGGGTGGCAAGGAGGTCCCTACCACTTGGAGAGAACCAT 2842
Db 518 lyProAlaGlyAsp---ArgGlyGlnAspGlyGlyProGlyLeuProGlyMetArgGlyL 537
QY 2841 TCCTGGG---CCTGGAAGCCAGATGTCAGGGGTTAAGNAGTGAGTGGGTGGAGAGAA 2785
Db 537 euProGlyLeuProGlySerPro-----GlyS 546
QY 2784 GTGGAGGACAGAGGCTGTAGACCACTCTTTCGAGAAATTTGGCAGGAGAAACCCAGGAGAGA 2725
Db 546 erAspGlyLysProGlyProPro-----GlyAsn-----GlnGlyGluProGly--- 560
QY 2724 AAAAGGTCGGGAACATAGAGGGCAGTGGAGTGAAGCAAAAGTTTCTGCACAAGGAAC 2665
Db 560 ----- 560
QY 2664 AAGAGAGATGAGAGTGGCCCATTTGGAACAGAGCCCATTTGCGG----- 2618
Db 561 -----ArgSerGlyPro-----ProGlyProAlaGlyProArGlyGlnProG 575
QY 2617 -----CTGACACACCCAGGCCACCCAGCT---GAGGGGCTG---GGGAGAGACGGGTACT 2569
Db 575 lyValMetGlyPheProGlyProLysGlyAsnGlyAlaProGlyLysAsnGlyGlu- 594
QY 2568 CACTAACACTCGAAGAGGGTTCAGCTGATAGGAGCGGTTCGAGAGAGGAGAAATG 2509
Db 595 -----ArgGlyProGlyGlyProProGlyThrProGlyProAlaGlyLysAsnG 611
QY 2508 GAGGCACTCATGATAGTACAGGAGGAGGCTATGGAAGCAGCATGTTATAGGGAAG 2449
Db 611 lyAsp-ValGlyLeuProGlyProPro-----GlyProAlaGlyProAlaGlyAsp 627
QY 2448 AGGCGAGAGCCCTTCCAAACCTCGGGGAGCGGCTGGGCGCCAGGAGGAGGATGAGGG 2389
Db 628 ArgGlyGluProGlyProSer-----GlySerProGlyLeuGlnGlyLeuProGlyGly 645
QY 2388 AGATGTATACTCGGAAGAGGAGGAGCAAGGGAGGAGGCTGGGATAGGCAAGGGCC 2329
Db 646 ProGlyProAlaGlyLysAsnGlyLysProGlyGluProGly----- 659
QY 2328 CAGGGGAGGTTCCACGCTGACAAGCAGACAGACAACTAATTTCTCGGTATCTCTG 2269
Db 660 -----ProLysGlyAspIle 664
QY 2268 GGGCTGGAGCAGCTCCCGCCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2209
Db 665 GlyGlyProGlyPheProGlyProLysGlyGluAsnGlyPro----- 679
QY 2208 GTCTGGGATGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2152
Db 680 -----GlyGluArgGlyProGlnGlyProGlyProGlyProThrGlyAlaArg 694
QY 2151 GCTCTCCAAACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2098
Db 695 GlyGly-----ProGlyProAlaGlySerGluGlyAlaLysGlyProProGly 710
QY 2097 CCCCCAATCTTACAGATACCTCCATCCACACCCCTAAACCCCTGCCACTACTGCTCCT 2038
Db 711 ProProGlyAlaProGlyGlyThrGlyLeuProGlyLeuGlnGlyMetPro-GlyGluAr 730
QY 2037 GGGTTCCTACAGTCCCTTCCCTCCATGACCCCTAGGAGGCCAGCTAGGATGGAGA 1978
Db 730 gGlyAlaSerGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyLysGly-- 749
QY 1977 GGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1918
Db 750 -AlaAspGly-----LeuProGlyAlaArg-----GlyGluAr 760
QY 1917 CGGAGCTGGAACGAGGAGGATGGGAGGTGGGGCCCTCTCTCCCAACCCCAACCCCGCCG 1858

RA VARIANTS AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
RA Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5
RT (IV) chain associated with X-linked Alport syndrome: characterization
RT of the mutation by direct sequencing of PCR-amplified lymphoblast
RT cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
[12]
RN VARIANTS AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
RA Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129(1992).
[13]
RN VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemlink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Cost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of
RT patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
[14]
RN VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
[15]
RN VARIANTS AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Pain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
[16]
RN VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
[17]
RN VARIANTS AS, AND VARIANTS ASP-430; SER-444; ASN-664 AND
RX MET-1428.
RA Knebelmann B., Brillat C., Forestier L., Arrondel C., Jacassier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
[18]
RN VARIANTS AS ASP-1498.
RX MEDLINE=96213932; PubMed=8829632;
RA Tverskaya S., Bobrykina V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of Al498D in noncollagen domain of a5(IV) collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
[19]
RN VARIANTS AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
[20]
RN VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
[21]
RN VARIANTS AS V-420; 456-P-P-458 DEL; D-573; D-624; D-635; 802-G-P-807
RX DEL; R-869; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357
AND R-1649.
RA MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
[22]
RN VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
RX SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RA MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
RT sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
[23]
RN VARIANTS AS ARG-822.
RX MEDLINE=20025011; PubMed=10563487;
Alignment Scores:
Pred. No.: 1,81e-07 Length: 1685
Score: 316.00 Matches: 336
Percent Similarity: 28.19% Conservative: 79
Best Local Similarity: 22.83% Mismatches: 498
Query Match: 3.35% Indels: 560
DB: 1 Gaps: 76
US-09-931-704-3 (1-5087) x CA54_HUMAN (1-1685)
Qy 1224 CTTCCATCCCTCTGCTCCCTCCATCTGTCCTCCACCTCACCCTCCACCCAGCCCAAGGT 1283
Db 268 ProProGlyPro---ProGlyIleArgGlyPro-----ProGlyPro---ProGly 282
Qy 1284 GGGACACACACCTGAGGGGTGCCAGCTGCTTCCC-----CGTGTGGCCCGGCGCGC 1337
Db 283 GlyGluLysGlyGluLysGlyGluGluGluProGlyLysArgGlyLysProGlyLys 302
Qy 1338 GCTCATGCTTCGTCCTCCATCTGTCACAGGAGGACTCGTGGGGAGTTAGCTGCTGT 1397
Db 303 AspGlyGluAsnGly-GlnProGlyIleProGlyLeuProGlyAspProGlyTyrPro-- 321
Qy 1398 GCACGGTCTCTGGCACCTCCTCGCAGTCCAGCTCTCAATCGCACGGGACCCAGGC 1457
Db 322 -----GlyGluProGlyArgAspGlyGluLysGlyGlnLysGlyAspThrGl 337
Qy 1458 CTGGCCCTCCATCCAGAAACCTATGACCTCACCGG----- 1494
Db 337 yProProGlyProGlyLeuValIleProArgProGlyThrGlyIleThrIleGlyGl 357
Qy 1495 -----CTACCTGGAGACCACTCGCAGCTTGGCTGGGACCTATGTA 1538
Db 357 uLysGlyAsnIleGlyLeuProGlyLeuProGly----- 368

QY 1539 GTATCCAGCGTAGGAATCTGGAGTTGGGA-----GGAGTGAGGAGT----- 1581
Db 369 -----GluysGlyGluatgGlyPheProGlyleGlnGlyProProGlyLe 384
QY 1582 -----TGGGAAGACAGCTCCTAACCGTGAGAGGTTCTGTAATGATGGGTGA 1631
Db 384 uProGlyProProGlyAlaAlaValMetGlyProProGlyProProGlyPheProGlyG 404
QY 1632 GGAGGGCTCTTTGGCTCCACCGAGTCCCTGTCTGTCTATCTCTCCCTCCCTCT 1691
Db 404 u-ArGlyGlnLysGlyAspGluGlyProProGlyleSerileProGlyProProGlyL 424
QY 1692 TAGGTGGCCCCCCTTCCCATCCCTG-----GCCCCAGGACTAGGCATGTGGCGAG 1745
Db 424 euAspGlyGlnProGlyAlaProGlyLeuProGlyProProGly----- 438
QY 1746 GCCTCGCACCCGCTTGGCCCATTCCTCCCACTGGC-----TGCCAGCCAGCCGCC 1796
Db 439 -----ProAladlyProHisileProProSerAspGluileCysGluProGlyProP 456
QY 1797 CGCCTCCCTGGGGCGGGAGTCTCTCTGTATTACACCGTGTGTGTGTCTCTTG 1856
Db 456 roGlyProProGlySerProGlyAsp----- 464
QY 1857 CGCGGGCGGGTGGTGGGACAGAGG----- 1885
Db 465 -----LysGlyLeuGlnGlyGlnGlyValLysGlyAspLysGlyAspThrCysPheA 483
QY 1886 -----GCCCACTCCCATCCCTCGTCTGACCCAGGGCTCCCTCGTCTGCTTCC 1922
Db 483 snCysileGlyThrGlyleSerGlyProProGlyGlnProGlyLeu----- 498
QY 1923 CCCCAGAGCTGGGGCCCTGCTGCTGACCCAGGGCTCCCTCGTCTGCTGCTTCC 1982
Db 499 -----ProGly-----LeuProGlyProProGlySerLeuGlyPheP 511
QY 1983 CATCTAGCTGGCTCTCTAGGG-----GGGTATNGGGG---AAGGGAGCTGAGGA 2033
Db 511 roGlyGlnLysGlyGluLysGlyGlnAlaGlyAlaThrGlyProLysGlyLeuProGlyI 531
QY 2034 ACCAGGAGTAGTGGCAGG---GGGTTAGGGTGTGATGAGGTTATGCTGTAAAGAT 2090
Db 531 leProGlyAlaProGlyAlaProGlyPhePro-----Glys 543
QY 2091 TTGGGGGTGGTCCA---GAGGTGTTTCAGAGAGCCAGGAGAGAGAGGAGG-----G 2141
Db 543 erLysGlyCluProGlyAspilleLeuThrPheProGlyMetLysGlyAspLysGlyL 563
QY 2142 TTGGAGAGCCGAGCA-----CCATGGGGAACCGGCCCTCTCTTCC 2183
Db 563 euGlySerProGlyAlaProGlyLeuProGlyLeuPro-GlyThrProGlyGlnAspGly 582
QY 2184 CGTGTCTCTTCCATCCAGACCTACTCTGAGCCAGGAGAGAGAGAGAGAGAG 2243
Db 583 LeuProGlyLeuProGlyProLysGlyGluProGlyleThrPheLysGlyGluArg 602
QY 2244 -----GTGGCGGGGAGCTGGCTCA-----GCC 2267
Db 603 GlyProProGlyAsnProGlyLeuProGlyLeuProGlyAsnileGlyProMetGlyPro 622
QY 2268 CCAGGATACCCGAGGAATAGTTGTCTGTCTGTCTGTCTGTCAGCGGTGAACCTCCCT 2327
Db 623 ProGlyPhe-----GlyProPro 628
QY 2328 GGGCCCTTGCCTATCCAGGCTCTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2387
Db 629 GlyProValGlyGluLysGlyleGlnGlyValAlaGlyAsnProGlyGlnProGlyle 648
QY 2388 TCCCTCATCCCTTCTCCCTGGGGCC-----CAGCGGCTC 2420
Db 649 -----ProGlyProLysGlyAspProGlyGlnThrIleThrGlnProGly 663
QY 2421 CCCCAGGGGTG-----GAAAGG 2438

Db 664 LysProGlyLeuProGlyAsnProGlyArgAspGlyAspValGlyLeuProGlyAspPro 683
QY 2439 GCTGTGGCTCTTCCCTATACCATGCTGTCTTCCATACGCTTCTCTCTCTCTCTCATG 2498
Db 684 GlyLeuProGlyGln-----ProGlyLeuProGlyleProGlySer 697
QY 2499 AGACTGCTCCATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2558
Db 698 LysGlyGluProGlyleProGlyleGlyLeu-----ProGlyProPro 712
QY 2559 AGTGTAGTAGTACCGCTCTCTCCAGCCCTCAGCTGCTGGCTGGCTGGCTGGCTGCTCAG 2618
Db 713 GlyProLysGlyPheProGlyleProGlyPro-----ProGlyAlaProGly---ThrPro 730
QY 2619 GGCAAAATGGGGC-----TCTGTTCCTTCCAAATGGGGCTCTCTCATCT 2657
Db 731 GlyArgileGlyLeuGluGlyProProGlyProProGlyPheProGlyProLysGlyGlu 750
QY 2658 CTCTCTTGTCTTGTGTGAGAAACCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2717
Db 751 ProGlyPheAlaLeuProGlyProPro-----GlyProProGlyLeuProGlyPheLys 768
QY 2718 CCCTTTTCTCTCTCT-----GGCTTCTCTCCCAATTTCTCTCA-----AGG 2759
Db 769 GlyAlaLeuGlyProLysGlyAspArgGlyPheProGlyProProGlyProProGlyArg 788
QY 2760 AGTGTCTTACACCTCTCTCT-----CCACTT----- 2786
Db 789 ThrGlyLeuAspGlyLeuProGlyProLysGlyAspValGlyProAsnGlyGlnProGly 808
QY 2787 CTCTTCCACCCACTCTTCTTAAACCCCTTCAATCTGGCTTCCAGGCCCCAGCAATGT 2846
Db 809 ProMetGlyPro-----ProGlyLeuProGlyle 818
QY 2847 TCTCTCAGGTGCTCAGGACCTCTTGGCCAGCCGACAGTGTTTGAAGGCTCATTC 2906
Db 819 GlyValGlnGlyProProGlyProPro-----GlyleProGlyProle 833
QY 2907 TCCTTGTCTGTCTTGTGAGCCACACTGCTGAGCGTGTCTCTCTCTCTCTCTCTCT 2966
Db 834 Gly-----GlnProGlyLeuHisGlyle 841
QY 2967 CTTTGTCTCTGCACTCTCTGGCCACCTTCTTACTCT-----CCAGCTCTCTCAGGCTCC 3023
Db 842 ProGlyGluLysGlyAspProGlyProProGlyLeuAspValProGlyProProGlyGlu 861
QY 3024 TCTTCTCTCTGTCTGCTGCCCCCAGCGGGCACTCTCCCAAGGTTTGGCCACCCAGCCAA 3083
Db 862 ArgGlySerProGlylePro-----GlyAlaProGlyProleGlyProProGlySer 879
QY 3084 TCAGCACGCTCTCTCTGAGCGTCTTGTGG----- 3113
Db 880 ProGlyLeuProGlyLysAlaGlyAlaSerGlyPheProGlyThrLysGlyGluMetGly 899
QY 3114 -----TCTCTCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3158
Db 900 MetMetGlyProProGlyProProGlyProLeuGlylePro----- 913
QY 3159 CACCCCACTGCTTCACTGTCTGATCAATATATCTTCTTCTTCTTCTTCTTCTTCT 3218
Db 913 ----- 913
QY 3219 GGAGGCCATGAACAAAGAGCCCTAGCATGGAGAGCGGCTGTCTCAGGGGACACAAAA 3278
Db 914 -----GlyArgSerGly 917
QY 3279 ATAGAACTTTGGGAGCAGGTATCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3338
Db 918 ValProGlyLeuLysGlyAspGlyLeuGlnGlyGlnProGlyLeuProGlyProThr 937
QY 3339 TCCCATCACCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3398

Db 938 GlyGluLysGlySerLysGlyGlu-ProGlyLeuProGlyProPro-----GlyPr 954
 QY 3399 ACTTCAACCTCCCGCTGG-----GGCAGAGACTCTGCCAGAGCCACTGTTGACT 3452
 Db 954 oMetAspProAsnLeuGluSerLysGlyGluLysGlyGluProGlyLeuProGlyLyl 974
 QY 3453 TGGAGGTGGCGAAGCTCAATGACAAACTGCG-----GCTGACCCAGAACTACGAGG 3506
 Db 974 eProGlyValSerGlyProLysGlyTyrglnGlnLeuProGlyAspProGlyGlnProGl 994
 QY 3507 CCTA-----CAGCCACTTCTGTGTACTGTGGCTCAACCGTCAGCTGCCACTG 3560
 Db 994 yLeuSerGlyGlnPro-----GlyLeuProGlyProProGly----- 1006
 QY 3561 CTGAGCTGCCCGCAGCTGCCCACTTCTGCACAGCTCCAGGCTCTGGGCTGCTGGGC----- 3616
 Db 1007 -----ProLysGlyAsnProGlyLeuProGlyGlnPro-GlyLeuLeuGlyProp 1023
 QY 3617 --AGCATTTGGGGCTCATGAGCTCTGGGCTACCCACTGCCCGCTGCTGGGA 3674
 Db 1023 roGlyLeuLysGlyThrIleGlyAspMetGlyPheProGlyProGlnGlyValGluGlyP 1043
 QY 3675 CTGAA---CCCACTTGAGCTCTGGCTGCCCTGCCACAGTACTTCTC-----CAGAAGA 3725
 Db 1043 roProGlyProSerGlyValProGlyGlnProGlySerProGlyLeuProGlyGlnLysG 1063
 QY 3726 TGGACGACTTCTGCTGCTGAAGAGCTGCAGACCTGGCTGGCTCGGCCAAGGACT 3785
 Db 1063 lyAsp----- 1064
 QY 3786 TCAACGGCTCAAGAGAAGATGAGCTCCAGAGCTGCAGTCAACCTG----- 3835
 Db 1065 -----LysGlyAspProGlyIleSerSerIleGlyLeuProGlyLeuP 1079
 QY 3836 -----CACCTGGGGCTCATGGCTTCTGACTTC 3863
 Db 1079 roGlyProLysGlyGluProGlyLeuProGlyTyProGlyAsnProGlyIleLysGlyS 1099
 QY 3864 TGACCTTCTCTCTTCTGCTCCCTCCCTTCAAAACCTGCTCCCACTTTGTGAGAGCCAGCCCT 3923
 Db 1099 erValGlyAspProGlyLeuProGlyLeuProGlyThrProGlyAlaLysGlyGlnProG 1119
 QY 3924 GTATGCCA-----ACACTTGTGAGCCAGGAGACAGAAAGTGTGAGCTCTGG 3971
 Db 1119 lyLeuProGlyPheProGlyThrProGlyProProGlyProLysGly----- 1134
 QY 3972 CCCTTTCTGGACCGCTGGCGGTGTGATGCGATCAGCCCTGTCTCTCCCTCCACTCCCA 4031
 Db 1135 -----IleSerGlyProProGlyA 1141
 QY 4032 AA---GGTCTACCGAGC-----TGGGAGGAGGTACAGTAGGC---CCTG 4070
 Db 1141 snProGlyLeuProGlyGluProGlyProValGlyGlyGlyGlyHisProGlyGlnPro- 1160
 QY 4071 TCTGTCTCTTTCTACAGGAATCATGCTCGAGGAGTGTGAAGTGTTCAGGTGTGTG 4130
 Db 1161 GlyProProGlyGluLysGlyLysProGlyGlnAspGlyIle----- 1174
 QY 4131 CAGAGGCGCTATGGCTCTCTCTTCTGCTTACCTTGGCCAGTGC-----CCACCC 4184
 Db 1175 -----ProGly-----ProLysGlyGlnLysGlyGluProGly 1185
 QY 4185 AGCCCTCAGGTGGCACTCTGGAGGCGAGGGGTGTGAGGGGCCACACACACATGCTT 4244
 Db 1186 GlnProGlyPheGlyAsnProGlyProProGlyLeuProGly-----Leu 1200
 QY 4245 TCTGGGGTGAAG---CCCTTTGGCTGCCCACTCTCTTGGATGGGTGTGCTCCCTTAT 4301
 Db 1201 SerGlyGlnLysGlyAspGlyGlyLeuProGlyIleProGlyAsnProGlyLeuProGly 1220
 QY 4302 CCCCNAATCACTCTATACATCCAATTTCAGGAACAAACATGGTGGCAATCTACACAAA 4361
 Db 1221 Pro----- 1221

QY 4362 AGAGATGAGATTACAGTGCAGGTTGGGTCTGTCATTGGAGGTGCCCTATAACACAGAA 4421
 Db 1221 ----- 1221
 QY 4422 GAGAAATACTGAAAGCACAGGGCGAGGACAGACACAGACCCAGGAGTCTCCAAA 4481
 Db 1222 -----LysGlyGluProGlyPheHisGlyPheProGlyVal----- 1233
 QY 4482 GCACAGAGTGGCAAAACACCGAGCTAGCATCAGGACCTTGCCTCGAATTGTCTCTCC 4541
 Db 1234 -----GlnGlyProProGlyProGlySerProGlyProAlaLeuGluGlyProLys 1251
 QY 4542 AGTATTAGGTCCTCTCTCTGCCCTTCCAGGCTATCTGGTGGTTCAGGCTGG 4601
 Db 1252 GlyAsnProGlyProGlnGlyProPro-----GlyArgProGlyLeuProGlyPro 1268
 QY 4602 GGAGGCAACCATACCCACACACAGGATTTCTCTGAAAGTTTACAATGCAGTAGCATTTT 4661
 Db 1269 GluGlyProPro----- 1272
 QY 4662 GGGGTGTAGGTGGCAGCTCCCTCC-----AAGGCCCTGCCCTCCAGCCACCCAC 4711
 Db 1273 GlyLeuProGlyAsnGlyGlyIleLysGlyGlyLysGlyAsnProGlyGlnPro----- 1290
 QY 4712 TCATGACTTAAGTGTGTGTTGTTATTAATATTATTATTTGGAGATGTTATTATTAGATG 4771
 Db 1290 ----- 1290
 QY 4772 ATATTATTGCAGAATTCTTATTCTGTATTAAACAATAAATGCTTGCCCCAGAACTTA 4831
 Db 1291 -----GlyLeu 1292
 QY 4832 GTCTCTTTGCCAGCTCACC-----CCTCTCTGGTCTCATCAGACTCTTGCC 4879
 Db 1293 ProGlyLeuProGlyLeuLysGlyAspGlnGlyProProGlyLeuGlnGlyAsnProGly 1312
 QY 4880 ACCCTGGCTCCACTCCCTCTGCTGCTGCTGGTGGAGCTGCACAGAGCTCTGGG----- 4933
 Db 1313 ArgProGlyLeuAsnGlyMetLysGlyAspProGlyLeuProGlyValProGlyPhePro 1332
 QY 4934 -----AAGAGCCCTCTCTCTCCCGCACTGGGCGATGGGCGCACTCAGACTTACCC 4987
 Db 1333 GlyMetLysGlyProSerGlyValPro-----GlySer-AlaGlyProGluGlyGluPr 1350
 QY 4988 ACTGTCTGTCCACCAACCAACCCCTTGATCCCTCAGTCTCCCTCCACACAGCTTCTGTCCAC 5047
 Db 1350 oGlyLeuIleGlyProProGlyPro----- 1358
 QY 5048 CCCAGTTTCCCTCACCCTCCACC 5069
 Db 1359 -ProGlyLeuProGlyProSer 1365

RESULT 7

ID EXTN TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
 GN HRGPNT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Xanthi; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich

RT glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1645(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 CC MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X13885; CAA32090.1; -;
 DR PIR; S06733; S06733.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Alignment Scores:

Pred. No.: 2.27e-07 Length: 620
 Score: 314.50 Matches: 217
 Percent Similarity: 30.36% Conservative: 42
 Best Local Similarity: 25.44% Mismatches: 268
 Query Match: 3.33% Indels: 326
 DB: 1 Gaps: 44

US-09-931-704-3 (1-5087) x EXTN_TOBAC (1-620)

Qy 1658 CCCCTGTCTGTCTATCTCTCCCTCCCTCTCTAGT---GGCCCCCCTCTCCCA 1714
 Db 36 ProProValThrSerGlnProProSerSerIleGlyLeuSerProSerAlaPro 55
 Qy 1715 TCCCTGGCCCCAGGACTAGGCATGTGGGAGCCCTCGCACCCGCTGGCCCATGGCCC 1774
 Db 56 ThrThrThrProProSer-----ArgGlyHisValProSerProArgHisAlaPro 72
 Qy 1775 ACTGGCTCCAGCCAGCCGCGCTCCCTCGGGGGGGGAGTCTCCTCTGTTA 1834
 Db 73 -----ProArgHisAlaTyrProProProSerHisGlyHisLeu----- 85
 Qy 1835 CACCGTGTGTGTCTCTTGGGGGGGGGTGGTGGG----- 1876
 Db 86 -----ProProSerValGlyGlyProProProHisArgGly 97
 Qy 1877 -----GACAGAGG---GCCCACTCCCATCCCTCGGTCCAGC---TCGCC 1918
 Db 98 HisLeuProProSerArgGlyPheAsnProProProSerProValIle-SerProSerHis 117
 Qy 1919 TCTGCCCCCAGACCTGGGG-----CCCTGCTGTCTGGACCCAGGGGCTCCCTCCGT 1972
 Db 117 sProProProSerTyrGlyAlaProProProSerHisGlyProGlyHisLeuProSer--- 136
 Qy 1973 CTGCCTCTCCCATCTAGTGGGCTCTCTAGGGGGGTCTATGGGGGAAGGAGCTGAGG 2032
 Db 137 -----HisGlyGlnArg----- 140
 Qy 2033 AACCCAGGCGTAGTGGCAGGGGGTTTAGGGTGTGGATGAGGTTATGTGAAGATT 2092
 Db 368 rLeuProProProProSerSerPro-----ProProProse 381
 Db 140 ----- 140
 Qy 2093 GGGGTGGTCCAGAGGTGTTTCAGAGAGCCAGGAGAG-----AAGGAAGGAGG 2140
 Db 141 -----ProProSerProSerHisGlyHisAlaProProSerGlyGly 154
 Qy 2141 GTTGGAGAGCCAGGAGCCATGGGAACCGGCCCTCTTCCCGTGTCTCTCTTCCA-- 2198
 Db 154 yHisThrProProArg-----GlyGlnHisProProSerHisArgArgProSerProPr 172
 Qy 2199 -----CATCCAGACCTACTCTGGAGCCAGGAGAAAGAAAGGAGAGGTGGCGG 2251
 Db 172 oSerArgHisGlyHisPro----- 178
 Qy 2252 GGAGCTGGCTCCACCCCGGATACACCGAGGAAATAGTTTGTCTGTCTGTCTGTCAGC 2311
 Db 179 -----ProProProThrTyr----- 183
 Qy 2312 GTGTGAACCTCCCTCCCTGGGCGCTTG-----CCTATCCAGGCGCTCTCC-- 2359
 Db 184 -AlaGlnProProProThrProIleTyrSerProSerProGlnValGlnProProProTh 203
 Qy 2360 CTTCTCCCTCTTCCAGTTTATATCATCTCCCTCATCCCTTTCCT-----GGGCC 2410
 Db 203 rTyrSerProProProProThrHisValGlnProThrProProSerProSerArgGlyHis 223
 Qy 2411 CCAGCCGCTCCCGGAGGTGGAAAGGCTCTGCCCTCTTCCCTATATACATGTGTCTT 2470
 Db 223 sGlnProGlnProProThrHisArgHisAlaProProThrHisArgHisAlaProProTh 243
 Qy 2471 CCATAGCCTCTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2530
 Db 243 rHisGlnProSerPro-----LeuArgHisLeuProPro----- 254
 Qy 2531 GCTCTATCAGCTGAACCTCTTCTGGAGGTGTAGTGAGTACCGTCTCTCTCCAGCC 2590
 Db 255 -----SerProArgArgGlnProGlnProProThrTy 265
 Qy 2591 CTCAGCTGGTGGCTGGGTGTGTGTCAGCGCAAAATGGGGCTCTGTTCATATGGGCCACT 2650
 Db 265 rSerProProProProAlaTyrAlaGlnSerProGlnProSerProThrTyrSerPro-- 284
 Qy 2651 CTCATCTCTCTTGTCTTGTGCGAGAAACCTTGTCTCTCTCTCTCTCTCTCTCTCTAG 2710
 Db 285 -----ProProProThrTyrSerProProProProPro 295
 Qy 2711 TTCCCGACCTTTTCTCTCTCTGGCTTTCCTCGCCAAATTTCTCCAAGAGGTGTCTACA 2770
 Db 295 r-----ProIleTyrSerPro-----ProProProAlaTyrSe 306
 Qy 2771 CCTCTGCTCCACTTCTCTCCACCACTCACTTCTTAAACCCCTGCAATCTGGCTTCC 2830
 Db 306 rProSerProProProThrProThrProThrPheSerProProProProAlaTyrSerPr 326
 Qy 2831 AGGCCCGAGCAATGTTCTCTCCAAGTCTGTCTCAGGACCTCTCTGCAAGCCGACAGT 2890
 Db 326 oProProThr-----TyrSerProPro----- 333
 Qy 2891 TTTTGAAGGCTCATTTCT 2950
 Db 334 -----ProPr 335
 Qy 2951 TTCTCGAACT 3007
 Db 335 oThrTyrLeuProLeuProProSerSerProIleTyrSerProProProProValTyrSerPr 355
 Qy 3008 AGCTCTCTCAGGCT 3067
 Db 355 oProProProProSer-----TyrSerProProProProProProProProThTy 368
 Qy 3068 TTGCCCGACCCAGCAATCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3127
 Db 368 rLeuProProProProProSerSerPro-----ProProProse 381

Db	563	SerThrSerSerThrSerSerGlySerPheGluAlaSerProValGluProGlnLeuPro	582
Qy	1912	GCTCGCCTCTGCCCCCAGACCTGGGGCCCTGCTGCTCTGGACCAGCGGGGCTCCCTTCG	1971
Db	583	SerLysGluGlyProGluPro-----ProGluGluValProPro-----Pro	596
Qy	1972	TCCTGCTCTCCATCTAGCTGGCCCTCTAGGGGGTGCATGGGGAGAGGGACTGTAGG	2031
Db	597	ThrThrProProValProLysValGluProLysGly---AspGlyIleGlyProThrArg	615
Qy	2032	GAACCCAGGCACGTAGTGGCAGGGGTTTAGGGTGTGGATGGAGGTTATGCTGTAAGGATT	2091
Db	616	GlnProProSer-----GlnGlyLeu-----GlyTyrProLysTyrArgGln	628
Qy	2092	TGGGGTGGTCCAGAGGTGTTTCAGAGAGCCAG-----GAGAGAAAG	2133
Db	629	LysSerLeuProProArgPheGlnArgGlnGlnGlnGluGlnLeuLeuLysGlnGlnGln	648
Qy	2134	AAGGAGGGTGGAGGACCGAGCCACCATGGGGAACCGGCCCTCTT-----	2181
Db	649	GlnHisGlnTrpGlnGlnHis---GlnGlnGlySerAlaProProThrProValProPro	667
Qy	2182	-----CCGTGTCTCTTCCATCCAGACCTCTCTGGAGCCAGGAAA	2229
Db	668	SerProGlnProValThrLeuGlyAlaValProAlaProGlnAlaProProProPro	687
Qy	2230	GAAGAAG-----GGAGAAGTGGCGGGGAGCTGGCTCAGGCCCCAGGAT---	2274
Db	688	ProLysAlaLeuTyrProGlyAlaLeuGlyArgProProProMetProProMetAsnPhe	707
Qy	2275	ACACCGAGGAATTAGTTTGTCTCTGTCTGCTGTCAGCGGTGAACCTCCCC-----	2328
Db	708	AspProArgTyrMetMet-----IleProTyrVal	718
Qy	2329	GGCCCTTGCCATCCAGGCCTCTCCCTTGCTTCTCCCTTCTTCCAGTATATACATCT	2388
Db	719	AspProArgLeuLeuGlnGlyArgPro---ProLeuGluPheTyrProGlyValHis	737
Qy	2389	CCCTCATCCCTT-----TCCTGGGCCCCAGCGCTCCGCC---	2424
Db	738	ProSerGlyLeuValProArgGluArgSerAspSerLeuGlyLeuSerSerGluProPhe	757
Qy	2425	-----GAG	2427
Db	758	AspArgHisAlaProAlaMetLeuArgGlyThrProProValAspProLysLeu	777
Qy	2428	GGTTGGAAGGGCTCTGCCCTCTCCCTATACCATGCTCTTCATAGCCTTCCTCTG	2487
Db	778	AlaTrp-ValGlyAspValPheThrAlaThrPro---AlaGluProArgProLeuThrSe	796
Qy	2488	TCCTACTCATGAGACTGCC-----TCCATTTCTCTCTTCGCAACCTGC	2532
Db	796	rProLeuArgGlnAlaAlaAspGluAspAspLysGlyMetArgSerGluThrProProVa	816
Qy	2533	TCCTATCAGCTGAACCTTCTTCGGAGTGTAGTACCGCTCTCTCCCGCCCT	2592
Db	816	lPro-ProProProTyrLeuAlaSer-----TyrProGlyPheProGluAsnG	833
Qy	2593	CAGCTGTGGG-----CCTGGGTGTGTGAGCGCA	2622
Db	833	lyAlaProGlyProProIleSerArgPheProLeuGluGluProGlyProArgProLeuP	853
Qy	2623	AATGGGCTCTGGTCCCATGGCCATCTCATCTCTCTTGTTCCTTGTGCAGAAAC	2682
Db	853	roTrpProProGlySerAspGluValAlaLysIleGlnThrProProProLysLysGluP	873
Qy	2683	CTTTGCTTCACCTCCACTGCC-----CTCTCTAGTT	2712
Db	873	roProLysGluGluThrAlaGlnLeuThrGlyProGluAlaGlyArgLysLeuProAlaAs	893
Qy	2713	CCCGA-----	2717

QY 3286 CTTTGGAGCAGGTATCTCTTGGTGTGAGCCAGCGCTCT-----GCCCTC 3333
Db GlnGluArgGluAsnAla : : : : :
QY 1247 GlnGluArgGluAsnAla : : : : :AlaArgGlySerGluGlyLysProSerLeu 1262
QY 3334 CTCCTTCCCATCACCTCTCTCTTTTCCAGCTCAACTACCTGGGCC-----CCCTTT 3387
Db ThrLeuPro : : : : :AlaSerAlaProGlyProGluGluAlaLeu 1275
QY 3388 CAAGCAGCCAGACTTCAACCTCTCCCGCT-----GGG 3420
Db ThrThrValThrValAlaProAlaProProAlaProAlaAlaLysSerProAspLeuSer 1295
QY 3421 GGCAGAGACTGCTCCAGGGCCACTGTTGACTTGGAGTGGCGAGCCCTCAATGCAA 3480
Db AsnGlnAsnSerAspGln : : : : :AlaAsnGluGluTrpGlu 1307
QY 3481 ACTCGGCTGAC-----CCAGAACTACGAGCCCTACAG-----CCACT 3519
Db ThrAlaSerGluSerSerAspPheThrSerGluArgArgGlyAspLysGluAlaProPro 1327
QY 3520 TCTGTGTTACTT-----GCTGGCTCAACCGTCAGCTGCCACTGCTGAGCTCG 3570
Db ProValLeuLeuThrProLysAlaValGlyThrProGlyGlyGlyGlyGlyAlaVal 1347
QY 3571 CCG-----GGGCTA 3648
Db ProGlyIleSerAlaMetSerArgGlyAspLeuSerGlnArgAlaLysAspLeuSerLys 1367
QY 3574 -----CAGCTGGC---CCACTTTCACACAGCTCCAGGGCTGCT 3612
Db ArgSerPheSerSerGlnArgProGlyMetGluArgGlnAsnArgArgProGlyProGly 1387
QY 3613 GGGCAGCAATTCGGCGCTCAGGAGCTCT-----GGGCTA 3648
Db GlyLys---AlaGlySerSerGlySerSerGlyGlyGlyGlyGlyGlyProGly 1406
QY 3649 CCCACTGCCAGCGCTGCTGGGAGCTGAACCCACTTGGAGCTCTGCTGGCCCTGCCCA--- 3705
Db ArgThrGlyProGlyArgGly-----AspLysArgSerTrpProSerProLys 1422
QY 3706 ---CAGTGACTTCTCCAGAGAT-----GAGCAGCTTCTGGCTGCTGAGGAGCT 3753
Db AsnArgSerArgProProGluGluArgProProGlyLeuProLeuProProProPro 1442
QY 3754 GCAGACCTGGCTGTG---GCGCTCGGCCAA-----GGACTCAA 3789
Db SerSerSerAlaValPheArgLeuAspGlnValIleHisSerAsnProAlaGlyIleGln 1462
QY 3790 CCGGCTCAAGAGAGATGCGCTCCAGCAGCTGCGACTGCGACCTGCGACCTGGGGCTCA 3849
Db GlnAlaLeuAlaGln---LeuSerSerArgGlnGlySerValThrAlaProGlyGlyHis 1481
QY 3850 TG-----CCTTCT 3857
Db ProArgHisLysProGlyProProGlnAlaProGlnGlyProSerProArgProProThr 1501
QY 3858 GACTTCTGACCTTCTCTCTGCTCCCTCCCTTCAAACTGCTCCACTTTGTGAGAGCC 3917
Db ArgTyArgProGlnArgValAsnSerGlyLeuSerSerAspProHisPheGluGluPro 1521
QY 3918 AGCCCTGTATGCCAACACCTGTTGAGCCAGGAGACAGAGCTGTGAGCTCTGGCCCTTT 3977
Db GlyProMetValArgGlyValGlyGlyThrProArgAspSerAlaGlyValSerProPhe 1541
QY 3978 CTGAGCCGGCTGGGGT-----GTGATCGATCAGCC 4010
Db Pro---ProLysArgArgGluArgProProArgLysProGluLeuLeuGlnGluSer 1560
QY 4011 CTGTCTCTCCACCTCCCAAGGTCTACCGAGCTGGGGAGAGGTACAGTAGGCCCTG 4070
Db LeuProProProHisSerSerGlyPheLeuGlySerLysProGlyProGln 1580

QY 4071 TCCTGCTCTCTTCTACAGGAAGTCACTCTCGAGGGAGTGTGAAGTGGTTCAGGTG--- 4127
Db AlaGluSerArgAspThrGlyThrGluAlaLeuThrProHisIleTrpAsnArgLeuHis 1600
QY 4128 -----GTGCAGAGCCCTCATGGCCCTCTGCTCTTTCGCCTACCA----- 4166
Db ThrAlaThrSerArgLysSerTyArgProThrSerMetGluProTrpMetGluProLeu 1620
QY 4166 ----- 4166
Db SerProPheGluAspValAlaGlyThrGluMetSerGlnSerAspSerGlyValAspLeu 1640
QY 4167 -----CTTGGCCAGTGC----- 4178
Db SerGlyAspSerGlnValSerSerGlyProCysSerGlnArgSerSerProAspGlyGly 1660
QY 4179 -----CCACCAGCCCTCAGTGGCCACATCTGG-----AGG 4210
Db LeuLysGlyAlaAlaGluGlyProProLysArgProGlyGlySerProLeuAsnAla 1680
QY 4211 GCAGGGGTGAGGGGCCA-----CCACCACATGCTTTC 4246
Db ValProCysGluGlyProProGlySerGluProProArgArgProProAlaProHis 1700
QY 4247 TGGGTGAAGCCCTTGGCTGCCCTCTCTGATGGGTGTGCTCCCTTATCCCA 4306
Db AspGlyAspArgLysGluLeuProArgGluGlnProProGlyProGlyGlyThr 1720
QY 4307 AATCACTCTATACATCCAAATTCAGGAACAACATGTTGGCAATTCACAAAAAGA 4366
Db GluArgSerGlnArgThrAspArgGlyThrGluProGlyProLysArgProSerHisArg 1740
QY 4367 TGAGATTAACAGTGCAGGGTGGG-----GTCTGCATT 4399
Db ProGlyProProValGlnGlnPheGlyThrSerAspLysAspSerAspLeuArgLeuVal 1760
QY 4400 GGAGTGCCTTAAACAGAGAGAAATACTGAACAGCAGGGGAGGACACACAG 4459
Db GlyAspSerLeuLysAlaGluGlyLeuThrAlaSerValThrGluAlaIleProVal 1780
QY 4460 ACCAGA-----CCAGGAGTCTCCAAACACAGAGTGGCAACAAA--- 4501
Db SerArgAspTrpGluLeuLeuProSerAlaAlaSerAlaGluProGlnSerLysAsn 1800
QY 4502 -----CCGAGCTGAGCATC-----AGCACCTGCTCGA 4531
Db LeuAspSerGlyHisCysValProGluProSerSerSerGlyGlnArgLeuTrpProGlu 1820
QY 4532 ATTGTCTTCAGTATTCAGTGCCTCTTCTGCCCCCTTCCAGGGTATCTGGGT 4591
Db ValPheTyGlySerAlaGlyProSerSer-SerGlnIleSerGlyGlySerHisGlyLeu 1840
QY 4592 G-----CCAGCTG-----GGAGGCAACCATAGCCACAC---CACAG 4627
Db uSerIleThrSerLysGlnTrpArgLeuArgProGlyThrProSerLeuHisProTyArg 1860
QY 4628 GATTTCTGAAAGTTTACATGCTAGTACATTTTGGGTGTAGGTGGCAGCTCCCAAG 4687
Db 1860 G-----SerGlnProLeu 1864
QY 4688 GCCCTGCC-----CCAGCCCCCAGCTCATGCTCTAGTGTGTGTATTAATATT 4741
Db uTyLeuProProGlyProAlaProProSer-----AlaLeuLeuSerGln 1879
QY 4742 TATTTATTGGAGATGTTATTATTATTAGATGATATTTATTGCAGAAATTTCTATTCTGTAT 4801
Db yValAlaLeuLysGlyGlnPhe-LeuAsp----- 1888
QY 4802 TAACAAATAAATGCTGCCAGAACTAGTCTCTTTGCCAGC-----CTCACC 4852
Db --PheSerThrMetGlnAlaThrGluLeuGly-LysLeuProAlaGlyGlyValLeuTy 1907
QY 4853 CCTCTGTTGCTCATCAGACTCTTGGCCCTTGGCTCCCACTCCCTGCTGCTCT----- 4907

[illegible]

641	----	ArgGlyHis--ProGlyVal-----	644
3204	ATTGGAAAACTCAGGGAGGCCATGAACAAAGACCTAGCATGGAGACAGGCCAGTGT	3263	
647	-----	ProGlyHisProGlyValArgGlyProAspGlyLeuLysGlyGln--L	662
3264	CAGGGGACACAAAAATAGAAAC-----	TTTGGGAGCAGGTATCTCTTGGTGGTGAGC	3317
662	ysGlyAspThrIleSerCysAsnValThrTyProGlyArgHis-----	GlyPro	679
3318	CAGCGGCTCGCCCTCTCTCTCCCCCATCA-----	3347	
679	roGly--PheAspGlyProProGlyProLysGlyPheProGlyProGlnGlyAlaProGly	698	
3348	-----	CCCTCTCTCTTTTTCACACTCACTACC	3374
699	LeuSerGlySerAspGlyHisLysGlyArgProGlyThrPro--GlyThrAlaGluIlePr	718	
3375	TGGGGCCCCCTTTCAACGAGCCAGACTT-----	3402	
718	oGlyProProGlyPheArgGlyAspMetGlyAspProGlyPheGlyGlyGluLysGlySe	738	
3403	-----	CAACCTCCCGCTGGGGC-----	3423
738	rSerProValGlyProProGlyProProGlySerProGlyValAsnGlyGlnLysGlyI1	758	
3424	----	AGAGACTCTGCCAGGGCCACTGTTGACTTGA-----	3456
758	eProGlyAspProAlaPheGlyHis-----	LeuGlyProProGlyLysArgGlyLeuSe	776
3457	-----	GGTGTGGCGAAAGCCCAATGACAAACTGGCGTACCACCAACTACGAGG	3506
776	rGlyValProGlyIleLysGlyProArg-----	GlyAspProGlyCysProGl	792
3507	CCTACAGCCACTCTCTGTGTTACTTCGGTGG-----	3537	
792	YAlaGluGlyProAlaGlyIleProGlyPheLeuGlyLeuLysGlyProLysGlyArgGl	812	
3538	-----	CCTCAACCGTCAGGCTGCCACTGC-----	3561
812	uGlyHisAlaGlyPheProGlyValProGlyProProGlyHisSerCysGluArgGlyAl	832	
3562	-----	TGAGCTGCGCGC-----	3573
832	aProGlyIleProGlyGlnProGlyLeuProGlyTyProGlySerProGlyAlaProGl	852	
3574	-----	CAGCTGGCCCACTTCTGCAACGAGCTCCAGGCGCTGTGGGAG-----	3618
852	yGlyLysGlyGlnProGlyAspVal-----	GlyProProGlyProAlaGlyMetLysGl	870
3619	-----	CATTGGCGGCGTCAATGCAGCTCTGGGCTACCCACTGCCCC	3659
870	YLeuProGlyLeuProGlyArgProGlyAlaHisGlyProProGlyLeuProGlyIlePr	890	
3660	AGCCCTGCCTGGGACTGAAACCCACTTGACCTCTGGCGC-----	TG	3701
890	oGlyProPheGlyAsp-----	AspGlyLeuProGlyProProGlyProLysGl	906
3702	CCACAGTACTTCTCT-----	3717	
906	yProArgGlyLeuProGlyPheProGlyPheProGlyGluArgGlyLysProGlyAlaGl	926	
3718	-----	CCAGAAGA	3725
926	uGlyCysProGlyAlaLysGlyGluProGlyGluLysGlyMetSerGlyLeuProGlyAs	946	
3726	TGGACGACTTCTGGCTGCTGAAGGAGCT-----	CGACACTGGC	3764
946	pArgGlyLeuArgGlyAlaLysGlyAlaIleGlyProProGlyAspGluGlyGluMetal	966	
3765	TGTGGCGCTCGGCCAAGCACTTCAA-----	CC	3791
966	aIleIleSerGlnLysGlyThrProGlyGluProGlyProProGlyAspAspGlyPhePr	986	

RA Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).
RN [4]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=84297217; PubMed=6473103;
RA Aho S., Tate V.E., Boedtker H.;
RT "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen
gene.";
RL Nucleic Acids Res. 12:6117-6125(1984).
RN [5]
RP SEQUENCE OF 1-79 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
of the chicken pro alpha 1(I) collagen mRNA.";
RL Gene 56:71-78(1987).
RN [6]
RP SEQUENCE OF 78-92.
RC TISSUE=Skin;
RX MEDLINE=71115216; PubMed=5544653;
RA Higberger J.H., Kang A.H., Gross J.;
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
peptides from chick and rat skin collagens.";
RL Biochemistry 10:610-616(1971).
RN [7]
RP SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.
RX MEDLINE=82058081; PubMed=6272119;
RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;
RT "Structure of the pro alpha 2 (I) collagen gene.";
RL Nature 294:129-135(1981).
RN [8]
RP SEQUENCE OF 78-92.
RC TISSUE=Skin;
RX MEDLINE=70131186; PubMed=4313735;
RA Kang A.H., Gross J.;
RT "Amino acid sequence of cyanogen bromide peptides from the amino-
terminal region of chick skicollagen.";
RL Biochemistry 9:796-804(1970).
RN [9]
RP SEQUENCE OF 78-92 AND 415-448.
RC TISSUE=Skin;
RX MEDLINE=69285369; PubMed=5809220;
RA Kang A.H., Igarashi S., Gross J.;
RT "Characterization of the cyanogen bromide peptides from the alpha-2
chain of chick skin collagen.";
RL Biochemistry 8:3200-3204(1969).
RN [10]
RP SEQUENCE OF 78-92 AND 415-448.
RC TISSUE=Bone;
RX MEDLINE=69206882; PubMed=5785223;
RA Lane J.M., Miller E.J.;
RT "Isolation and characterization of the peptides derived from the
alpha 2 chain of chick bone collagen after cyanogen bromide
cleavage.";
RL Biochemistry 8:2134-2139(1969).
RN [11]
RP SEQUENCE OF 566-587 FROM N.A.
RX MEDLINE=79074929; PubMed=364479;
RA Lehrach H., Frischauf A.-M., Hanahan D., Wozney J., Fuller P.,
RA Kravchenkov R., Boedtker H., Doty P.;
RT "Construction and characterization of a 2.5-kilobase procollagen
clone.";
RL Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).
RN [12]
RP SEQUENCE OF 902-1362 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller P., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [13]
RP SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.
RX MEDLINE=81264246; PubMed=6267043;

RA Dickson L.A., Ninomiya Y., Bernard M.P., Pesciotta D.M., Parsons J.,
RA Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,
RA Fietzek P.P., Olsen B.R.;
RT "The exon/intron structure of the 3'-region of the pro alpha 2(I)
collagen gene.";
RL J. Biol. Chem. 256:8407-8415(1981).
RN [14]
RP SEQUENCE OF 932-954 AND 968-980 FROM N.A.
RX MEDLINE=81064671; PubMed=6159982;
RA Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,
RA de Crombrughe B.;
RT "Correlation between splicing sites within an intron and their
sequence complementarity with U1 RNA.";
RL Cell 21:689-696(1980).
RN [15]
RP SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.
RX MEDLINE=81112157; PubMed=7460017;
RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,
RA Itani M., Pastan I., de Crombrughe B.;
RT "The collagen gene: evidence for its evolutionary assembly by
amplification of a DNA segment containing an exon of 54 bp.";
RL Cell 22:887-892(1980).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
(FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25963; AAA69960.1; JOINED.
DR EMBL; M25956; AAA69960.1; JOINED.
DR EMBL; M25959; AAA69960.1; JOINED.
DR EMBL; M25961; AAA69960.1; JOINED.
DR EMBL; M25962; AAA69960.1; JOINED.
DR EMBL; M25965; AAA69961.1; JOINED.
DR EMBL; M25964; AAA69961.1; JOINED.
DR EMBL; M25984; AAA69962.1; JOINED.
DR EMBL; M25957; AAA69962.1; JOINED.
DR EMBL; M25966; AAA69962.1; JOINED.
DR EMBL; M25967; AAA69962.1; JOINED.
DR EMBL; M25969; AAA69962.1; JOINED.
DR EMBL; M25970; AAA69962.1; JOINED.
DR EMBL; M25971; AAA69962.1; JOINED.
DR EMBL; M25972; AAA69962.1; JOINED.
DR EMBL; M25973; AAA69962.1; JOINED.
DR EMBL; M25974; AAA69962.1; JOINED.
DR EMBL; M25976; AAA69962.1; JOINED.
DR EMBL; M25977; AAA69962.1; JOINED.
DR EMBL; M25978; AAA69962.1; JOINED.
DR EMBL; M25979; AAA69962.1; JOINED.
DR EMBL; M25980; AAA69962.1; JOINED.
DR EMBL; M25981; AAA69962.1; JOINED.
DR EMBL; M25982; AAA69962.1; JOINED.
DR EMBL; M25983; AAA69962.1; JOINED.
DR EMBL; J00826; AAA51611.1; JOINED.
DR EMBL; J00821; AAA51611.1; JOINED.
DR EMBL; K00792; AAA51611.1; JOINED.
DR EMBL; J00830; AAA51613.1; JOINED.
DR EMBL; J00829; AAA51613.1; JOINED.
DR EMBL; J00837; AAA51614.1; JOINED.
DR EMBL; J00812; AAA51615.1; JOINED.
DR EMBL; J00811; AAA51615.1; JOINED.
DR EMBL; J00814; AAA51615.1; JOINED.

DR EMBL; J00815; AAAS1615.1; JOINED.
 DR EMBL; X02657; CAA26493.1; -
 DR EMBL; K00794; -; NOT_ANNOTATED_CDS.
 DR EMBL; V00390; CAA23688.1; -
 DR EMBL; X00760; CAA25330.1; -
 DR EMBL; M17608; AAA48673.1; -
 DR EMBL; M10581; AAA48637.1; -
 DR EMBL; M10540; AAA48638.1; -
 DR EMBL; J00828; AAAS1612.1; -
 DR EMBL; J00827; AAAS1612.1; JOINED.
 DR EMBL; J00831; -; NOT_ANNOTATED_CDS.
 DR EMBL; J00832; -; NOT_ANNOTATED_CDS.
 DR EMBL; J00833; -; NOT_ANNOTATED_CDS.
 DR EMBL; J00822; -; NOT_ANNOTATED_CDS.
 DR PIR; A02868; CGCH28.
 DR PIR; S07354; S07354.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001670; Fe-ADH.
 DR InterPro; IPR000885; Fib collagen C.
 DR InterPro; IPR001431; Peptidase M16.
 DR InterPro; IPR000244; Ribosomal_L9.
 DR InterPro; IPR000914; SBP_bac_5.
 DR InterPro; IPR000360; Transketolase.

Alignment Scores:
 Pred. No.: 4.84e-07 Length: 1362
 Score: 306.50 Matches: 354
 Percent Similarity: 30.58% Conservative: 86
 Best Local Similarity: 24.60% Mismatches: 520
 Query Match: 3.28% Indels: 480
 DB: 1 Gaps: 72

US-09-931-704-3 (1-5087) x CA21_CHICK (1-1362)

Qy 4305 GGGGATAGGAGGAGCAACCCATCCAGAGAGTGGGGAGGCCAAAGGCTTCACCCAG 4246
 Db 37 GlyAspLysGlyPro-----GlnGlyGluArgGly--ProProGly---ProProGly 52
 Qy 4245 AAAGCATGTGTGGGNGGCCCTCAACCCCTGCCCTCCAGATGTGCCA---CCTGAGG 4189
 Db 52 LysAspGlyGluAspGlyPro-----ProGlyProProGlyProProGlyProProG 70
 Qy 4188 GGCTGGTGGGCACTGG----- 4172
 Db 70 LysGlyGlyAsnPheAlaGlnTyrAspProSerLysAlaAlaAspPheGlyProG 90
 Qy 4171 -----CCAA 4168
 Db 90 LysProMetGlyLeuMetGlyProArgGlyProProGlyProGlyAlaSerGlyProProGlyProP 110
 Qy 4167 GTGGTAGGCAAGAGCAGGAGCCATGAGCGCTCTGCACCAACCTGAACCACTTCACAC 4108
 Db 110 roGlyPheGlnGlyValProGlyGlu----- 118
 Qy 4107 TCCCTCGAGCATGACTTCTGTAGAAACAGACAGGAGCGGCTTACTGTACT----- 4054
 Db 119 -----ProGlyGluProGlyGln-ThrGlyProGlnGlyProArgGly 132
 Qy 4053 CCTCCCAGCTCGGTAGACCTTTGGAGGTGGGGAGGAGACAGGGCTGATCCATCACAC 3994
 Db 133 ProPro-----GlyProProGlyLysAlaGlyGluAspGlyHisProGlyLysPro 149
 Qy 3993 GCCCAGCGGTCCAGGAAAGGCCAGAGCTCACAGCTTCTGTCT-----CCTGGCTCA 3940
 Db 150 GlyArgProGlyGluArgGlyValAlaGlyProGlnGlyAlaArgGlyPheProGlyThr 169
 Qy 3939 ACAGGT---GTTGGCATACAGGCTGGCTCTCACAAAGTGGAG-CAGGTTTGAAGGGG 3884
 Db 170 ProGlyProProGlyPheLysGlyIleArgGlyHisAsnGlyLeuAspGlyLeuThrGly 189
 Qy 3883 GAGCCAGAGGAGAGGTTCAGAGTCAAGAGCCATGAGCCCGCCAGGTCGAGGCTGACTCC 3824
 Db 190 GlnProGlyAlaProGlyThrLysGlyGluProGlyAlaProGlyGluAsnGly-ThrPr 209

Qy 3823 AGCTGTGGAGGTGTCATCTTCTTTGAGCCGGTTGAAGTCTTGGCCGAGCCGACAG 3764
 Db 209 oGlyGlnProGlyAla-----ArgGlyLeu-ProGlyGluArg 222
 Qy 3763 CCAAGTCTGAGGCTCTTTCAGCAGCCAGAGTCTCTCATCTTCTGGAGGA----- 3714
 Db 222 LysGlnGlyAlaProGlyProAlaGlyAlaArgGlySerAspGlySerAlaGlyPro 242
 Qy 3713 -----AGTCACTGTGGGAGGCCAGGAGTCCAAGTGGGTTTCAGTCCCGAGGCGG 3662
 Db 242 hrGlyProAla***** 262
 Qy 3661 CTGGGCGAGTGGT-----AGCCCAGAGCTGCCATGACGCCCGCAATGCTGCCAGCAG 3608
 Db 262 *GlyGluIleGlyProAlaGlyAsnGlyPro-----ThrGlyProAlaG 278
 Qy 3607 GCCTGAGGCTGTGTC-----AGAAGTGGCCAGGCTGGCGGCAGCTCAGC 3560
 Db 278 LysProArgGlyGluIleGlyLeuProGlySerSerGlyProVal----- 292
 Qy 3559 AGTGGCAGCTGACGGTTGAGGCCAGCGCAAGTAACACAGAAGGTGGTGTAGGCTCGTA 3500
 Db 293 -----GlyProProGlyAsnProGlyAlaAsnGlyLeuProGlyAla 306
 Qy 3499 GTTCTGGGTGAGCGCAGTTTGTTCATTGAGGCTTCGCCACACCTCCAAAGTCAACAGTGGC 3440
 Db 306 LysGlyAlaAlaGlyLeuPro-----GlyValAlaGlyAlaProGlyLeuProGlyP 324
 Qy 3439 CCTGGGCGAGTCTCTGCCCGCCAGCGGGGAGGGTTGAAGTCTGGCTGCT---TGAAAGG 3383
 Db 324 roArgGlyIleProGlyProProGlyProAlaGlyProSerGlyAlaArgGlyLeuValG 344
 Qy 3382 GGGGCCAGGTTAGTTCAGCTGTGAAAGGAGAGGTGATGGGAGGAGGAGGAGGAGCAGAGC 3323
 Db 344 LysGlyProGlyProAlaGlyAla-LysGlyGluSerGlyAsnLysGly-GluProGlyAla 363
 Qy 3322 CGCTGGCTCACCAAGGAGATACCTGCTCCCAAGTTCTATTTTGTGTCCCT--- 3265
 Db 363 aAlaGlyProPro-----GlyProPr 370
 Qy 3264 -----GA 3263
 Db 370 oGlyProSerGlyGluGlyLysArgGlySerAsnGlyGluProGlySerAlaGlyPr 390
 Qy 3262 CACTGGCCCTGCTCTCCATGCTAGGCTTC---TTTGTTCATGGCCCTCCCT---GAGTTTTT 3209
 Db 390 oProGlyProAlaGlyLeuArgGlyGluProGlySerArgGlyLeuProGlyAlaAspG 410
 Qy 3208 CCAATAAGGATATCATTTGTATGACAGGTGACAGTTGAAGCAGTGGC-----GGTGGTGA 3155
 Db 410 YArgAlaGlyValMetGlyProAlaGlyAsnArgGlyAlaSerGlyProValGlyAlaL 430
 Qy 3154 GCTCTCCAATGAGGAGCGGTAGAAAAGAGGAG----- 3121
 Db 430 sGlyProAsnGlyAspAlaGlyArgProGlyGluProGlyLeuMetGlyProArg***** 450
 Qy 3120 -----GAGGAGAGCGCACAGCGCTCAGGAAGGACGTGCTGATTGGCTGGGTGGG 3071
 Db 450 ***** 470
 Qy 3070 CAAACCTTGGGAGATGCCCGCTGTGGG---GCAGGACAGAGAGAGAGAG----- 3023
 Db 470 YAlaAsp-GlyArgValGlyProIleGlyProAlaGlyAsnArgGly-GluProGlyAsn 489
 Qy 3022 -----GAGCCCTGGAGGAGCTGGAGAGGTAGAA 2996
 Db 490 lLeGlyPheProGlyProLysGlyProThrGlyGluProGlyLysProGlyGlu---Lys 508
 Qy 2995 GGTGGCCCGAGGAGTGCAGAGACCAAGAGAGGAGGATTCGAGAAGCAGCAGCGCTCAG 2936
 Db 509 GlyAsnValGlyLeuAlaGlyProArgGlyAlaProGlyProGlyGlyAsnAsnGlyAla 528

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).
 RC TISSUE=Pituitary; PubMed=10077563;
 RX MEDLINE=99178941; PubMed=10077563;
 RA Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
 RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644 (1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).
 RC TISSUE=Cervical adenocarcinoma;
 RX MEDLINE=9919215; PubMed=10097068;
 RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;
 RT "SMRT, a silencing mediator for retinoid and thyroid hormone
 receptors-extended isoform that is more related to the nuclear
 receptor corepressor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524 (1999).
 RN [3]
 RP SEQUENCE OF 1023-2517 FROM N.A.
 RC TISSUE=Cervical adenocarcinoma;
 RX MEDLINE=96008552; PubMed=7566127;
 RA Chen J.D., Evans R.M.;
 RT "A transcriptional corepressor that interacts with nuclear hormone
 receptors.";
 RL Nature 377:454-457 (1995).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM TRAC-1).
 RC TISSUE=Fetal liver;
 RX MEDLINE=96408715; PubMed=8813722;
 RA Sande S., Privalsky M.L.;
 RT "Identification of TRACs (T3 receptor-associating cofactors), a family
 of cofactors that associate with, and modulate the activity of,
 nuclear hormone receptors.";
 RL Mol. Endocrinol. 10:813-825 (1996).
 RN [5]
 RP SEQUENCE OF 428-613 FROM N.A.
 RC TISSUE=Brain cortex;
 RX MEDLINE=9736492; PubMed=9225980;
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
 RT "CDNAs with long CAG trinucleotide repeats from human brain.";
 RL Hum. Genet. 100:114-122 (1997).
 CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
 PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
 ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIB.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: SMRT/TRAC-2 (SHOWN HERE) AND
 TRAC-1; ARE PRODUCED BY ALTERNATIVE SPLICING. TRAC-1 CONTAINS ONLY
 THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN
 ANTIREPRESSOR.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS. HIGH LEVELS OF EXPRESSION ARE
 DETECTED IN LUNG, SPLEEN AND BRAIN.
 CC -!- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.
 CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
 INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 DOMAINS (ID1 AND ID2).
 CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
 SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
 FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 SPECIFICITY.
 CC -!- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 CORNR BOXES.
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS

CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF113003; AAD20946.1; -;
 DR EMBL; AF125672; AAC22973.1; -;
 DR EMBL; U37146; AAC50236.1; -;
 DR EMBL; S83390; AAB50847.1; -;
 DR EMBL; U80750; AAB91446.1; -;
 DR TRANSFAC; T04689; -;
 DR Genew; HGNC:7673; NCOR2.
 DR MIM; 600848; -;
 DR InterPro; IPR001005; Myb DNA binding.
 DR Pfam; PF00249; myb DNA-binding; 2.
 DR SMART; SM00395; SANT; 2.
 DR PROSITE; PS00909; MYB_3; 1.
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 KW Coiled coil; Alternative splicing.
 FT DOMAIN 174 215
 FT DOMAIN 254 312
 FT DNA_BIND 429 474
 FT DNA_BIND 613 657
 FT DOMAIN 522 561
 FT DOMAIN 778 820
 FT DOMAIN 2139 2143
 FT DOMAIN 2342 2346
 FT DOMAIN 494 510
 FT DOMAIN 682 685
 FT DOMAIN 994 1002
 FT DOMAIN 1384 1389
 FT DOMAIN 1842 1846
 FT DOMAIN 2479 2482
 FT VARSPLIC 1 1702
 FT VARSPLIC 2353 2398
 FT CONFLICT 295 295
 FT CONFLICT 309 309
 FT CONFLICT 352 352
 FT CONFLICT 365 365
 FT CONFLICT 612 613
 FT CONFLICT 711 711
 FT CONFLICT 724 740
 FT CONFLICT 787 796
 FT CONFLICT 804 804
 FT CONFLICT 814 814
 FT CONFLICT 817 817
 FT CONFLICT 889 889
 FT CONFLICT 1023 1030
 FT CONFLICT 1034 1034
 FT CONFLICT 1894 1894
 FT CONFLICT 2494 2494
 SQ SEQUENCE 2517 AA; 274031 MW; F5805C01761258C0 CRC64;
 Alignment Scores:
 Pred. No.: 4.62e-07 Length: 2517
 Score: 306.50 Matches: 420
 Percent Similarity: 30.52% Conservative: 148
 Best Local Similarity: 22.57% Mismatches: 600
 Query Match: 3.25% Indels: 693
 DB: 1 Gaps: 99
 US-09-931-704-3 (1-5087) x NCR2_HUMAN (1-2517)
 QY 37 GCTTCGCGGAGCGCGGCTC-----GCCCTCCCA 66
 Db 955 AlaserProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAlaIlePro 974

QY 67 -----CTCCGCCAGCCTCCGGAGAGGAG-----CCGACCCG 99
Db 975 ProIleGlnValThrLysValHisGluProProArgGluAspAlaAalaProThrLysPro 994
QY 100 GCCGGCCAGCCCGCCGACCTCCGAGAGGTTGAAAC-----CCTAACTAG 153
Db 995 AlaProAlaProProGlnAsnLeuGlnProGluSerAspAlaProGlnGln 1014
QY 154 CCCTGCTCTTCATACATGACAGCAGCCGCCCTCTGATACCTAAACCGCAAGTCAC 213
Db 1015 ProGlySerProArgGlyLysSerArg-----SerPro 1026
QY 214 AGCCCTCCA-----ACTCACCTCTGCTGCCAGACCTCACCA-----CAT 255
Db 1027 AlaProAlaAspLysGluAlaPheAlaGluAlaGlnLysLeuProGlyAspPro 1046
QY 256 CTTGTGGACTCAACCTCAACGCACTAAATCAACCAATCCCAAGTCTAACTAATCT 315
Db 1047 ProCysTrpThrSerGlyLeuProPhe-----ProValProPro-----Arg 1060
QY 316 GAAACTTTTAAAGTAACCCAGTCCTTAAACCTAACCTAGCCCAATGCCAATTATATCTAC 375
Db 1061 GluValLysAlaSer-----ProHisAla----- 1069
QY 376 CTTAGCCAAACCTTAACCTGCTTCCAGTCCAAAGTGTCCACTGAATCTCACTCGT 435
Db 1070 ProAspProSerAlaPheSerTyrAlaProProGlyHisProLeu-----ProLeuGly 1087
QY 436 CTTACTGAAATCCAGAAAGCATATTTCCCACTGCCCA----- 477
Db 1088 LeuHis-----AspThrAlaArgProValLeuProArgProProThrIleSerAsnProPro 1106
QY 478 -----CATCCCTCCTTA-----CAGCACCCCAACCTCGGCC 507
Db 1107 ProLeuLysSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSer 1126
QY 508 TCTGGACTCTGATCTCTGGATCTCCAACTGTGAGTGCATCAGCAACCAAGCCG 567
Db 1127 GlnGlyMet-----SerValGlnLeuHisValProTyrSer--G 1139
QY 568 ACTGCTCAATGCACCT-----CTCTCCCTCTCTGCCC----- 602
Db 1139 LuHisAlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProL 1159
QY 603 --ACCTTGCGAGCTGATGAAAGCGCTCATTTGAAGTCCAACTTTTCCCACTTAACACC 660
Db 1159 ysLysLeuAlaPro---PheSerGlyValLysGlnGlnLeuSerPro----- 1174
QY 661 AAGAACGGGGTAA-----CTCCACACTGCCCGTTCTGAGAGTGACACTAAT 714
Db 1175 -----ArgGlyGlnAlaGlyProPro-----GluSerLeuGlyValProThrAlaG 1190
QY 715 CTCCTTCAATCTAACCCCACTACCTTCCACACTCAGGAATCACATCTCTAGATATA 774
Db 1190 InGluAla-----SerValLeuArgGlyThrAlaLeu----- 1200
QY 775 CCAAACTAAGCCCAATAGCAGCCGACCTAGTGTCTAACTATACCTTGTCTTC 834
Db 1201 -----GlySerValProGlyGlySerIleThrLysGlyLe---ProSerThrArgValP 1218
QY 835 CTATGGTGAGTCTGTCTTTGGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 894
Db 1218 ro---SerAspSerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAsp- 1236
QY 895 CTGCTCAGCTGCCAGCTGTGATGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 954
Db 1237 ValLeuTyrLysGlyThrIleThrArgIleIleGlyGluAsp---SerProSerArgLeu 1255
QY 955 CAG-----TGAGACTGGAAGACTGCG 975
Db 1256 AspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGlnGlyLysGly 1275
QY 976 AGGAAGCTAGGTACAACTGGGAACACAGCAGGTCGACCTGCAGT----- 1020

Db 1276 HisValLeuSerTyrGluGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArg 1295
QY 1021 -----CCTAGGCTGGCCCGCTCCCTCCATGTACACACATATACATGTT 1065
Db 1296 SerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGlu 1315
QY 1066 GGCACACACAGTGGGACACATCCCAAAGACTCTCTCAGCTGACACACATCCATCTCT 1125
Db 1316 Gly-----ArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMetGlyArg 1333
QY 1126 CAATATCTACTATAGACACTCATGCTGCCAAGTCTCAT-----CCTCAACA 1176
Db 1334 AlaIleProProGluArgHis-----SerProHisHisLeuLysGluGlnHis 1349
QY 1177 TACATGCT 1230
Db 1350 HisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAsp 1369
QY 1231 -----CCCTCTGCTCCCATCTGTGTGTCTCCACCTCACTCCACCCACCC 1272
Db 1370 TyrLeuArgArgGluAlaLysLeuLysArgGluGlyThrProProProProPro 1389
QY 1273 CAGCCCAAGTGGGACACACACTGAGGGCTGCCAGCTGCTCCCGCTGGGCCCGG 1332
Db 1390 SerArgAspLeu-----ThrGluAlaTyrLysThrGlnAlaLeu 1402
QY 1333 GCCGGCTCTATCT 1392
Db 1403 GlyProLeuLysLeu-----LysProAlaHisGluGlyLeuValAlaThrValLys-- 1419
QY 1393 CTTGTGCAGGTGCTCTGCGACCTCTGAGTGCAGCTCTCAATCGCAGGGGACCC 1452
Db 1420 -----GluAlaGlyArgSerIleHisGluIlePro 1429
QY 1453 AGGGCTGGCCCTCCAT---CCAGAAAACCTATGACCTCACCCCTACCTGGAGCACCA 1509
Db 1430 ArgGluGluLeuArgHisThrProGlu-----LeuProLeuAlaPro 1443
QY 1510 ACTCGCAGCTTGGCTGGGACCTATGTAGTATCAGCGTATAGGAATCTGGGAGTTGGGA 1569
Db 1444 ArgProLeuLys-----GluGlySerIleThrGln 1453
QY 1570 GGAGTGGAGGTGGGAAAGACACTCTTAACCTGGAGGTTCTGGTAAATGATGGGT 1629
Db 1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLys-HisAspVa 1473
QY 1630 GAGGAGGGCTCTTTGGCTCCACACCTGCTCTGCTCTATCTCTCTCTCTCTCTCTCT 1689
Db 1473 IArg-----SerIle 1476
QY 1690 CTTAGTGGCCCT---CCCACTTCCCATCTCCCTGGCCCGCAGGACTAGGATGGGACAG 1746
Db 1476 uIleGlySerProGlyArgThrPheProProValHisPro---LeuAspValMetAlaAs 1495
QY 1747 CTTCCGACCCCGCTTTGGCCCATTTGCCACTGGCTGCGCCAGCCCGCCCGCTCCCC 1806
Db 1495 pAlaArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAl 1515
QY 1807 TGGGGCCGGGAAGTCTCTCTGTTTACACCTGTTGGGTGCTCTTTCGCGCGCGG 1866
Db 1515 aSerSerSerGlyGly-----SerIleAla----- 1523
QY 1867 GTTGGGTGGGACAGAGGGGCCCACTCCCATGCTGCTGCTTCCAGCTCGCTCTCCCC 1926
Db 1524 -----ArgGlyAlaProValIleValProGluLeu-----GlyLysPr 1536
QY 1927 CAGACTGGGGCCCTGTCTCTGTGGACCCAGGGGCTCTCTCTCTCTCTCTCTCTCTCT 1986
Db 1536 oArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProAr 1556
QY 1987 CTAGCTGGGCCCTCTAGGGGGTCTATGGGGGAGGGGACTGTAGGGAACCCAGGCAGTAG 2046

RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
 RT "Organization of the exons coding for pro alpha 1(II) collagen N-
 RT propeptide confirms a distinct evolutionary history of this domain of
 RT the fibrillar collagen genes.";
 RL Genomics 4:438-441(1989).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans.";
 RL PNAS 88: 5:2052-2056(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=92755959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP VARIANT SER-1074.
 RX MEDLINE=97255959; PubMed=2572591;
 RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
 RA Hollister D.W.;
 RT "Glycine to serine substitution in the triple helical domain of pro-
 RT alpha 1 (II) collagen results in a lethal perinatal form of short-
 RT limbed dwarfism.";
 RL J. Biol. Chem. 264:18265-18267(1989).
 RN [13]
 RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.
 RX MEDLINE=89286907; PubMed=2543071;
 RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
 RT "Identification of the molecular defect in a family with
 RT spondyloepiphyseal dysplasia.";
 RL Science 244:978-980(1989).
 RN [14]
 RP VARIANT OSTEOARTHRITIS CVS-650.
 RX MEDLINE=90370826; PubMed=1975693;
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 RT cause of primary osteoarthritis associated with a mild
 RT chondrodysplasia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
 RN [15]
 RP VARIANT OI-IV VAL-717.
 RX MEDLINE=91291136; PubMed=2064612;
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770(1991).
 RN [16]
 RP VARIANT OSTEOARTHRITIS CVS-650.
 RX MEDLINE=91086471; PubMed=1985108;
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;
 RT "Cartilage expression of a type II collagen mutation in an inherited
 RT form of osteoarthritis associated with a mild chondrodysplasia.";
 RL J. Clin. Invest. 87:357-361(1991).
 RN [17]
 RP VARIANT HYPOCHONDROGENESIS GLU-984.
 RX MEDLINE=93054548; PubMed=1429602;
 RA Bogart R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
 RA Cohn D.H., Eyre D.R.;
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
 RT 1(II) chain produces hypochondrogenesis.";
 RL J. Biol. Chem. 267:22522-22526(1992).
 RN [18]
 RP VARIANT HYPOCHONDROGENESIS SER-705.
 RX MEDLINE=92262484; PubMed=1374906;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee B.;
 RT "Characterization of a type II collagen gene (COL2A1) mutation

RT identified in cultured chondrocytes from human hypochondrogenesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
 RN [19]
 RP VARIANT WS-II ASP-198.
 RX MEDLINE=93304428; PubMed=8317498;
 RA Koerkoe J., Ritvaniemi P., Haataja L., Kaaerlahti H.,
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 RT for glycine alpha 1-67 and that causes cataracts and retinal
 RT detachment: evidence for molecular heterogeneity in the Wagner
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
 RL Am. J. Hum. Genet. 53:55-61(1993).
 RN [20]
 RP VARIANT SEDC CVS-840.
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
 RA Eyre D.R.;
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
 RL Am. J. Hum. Genet. 53:A209-A209(1993).
 RN [21]
 RP VARIANT OSTEOARTHRITIS CVS-650.
 RX MEDLINE=93282819; PubMed=8507190;
 RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;
 RT "Human cartilage from late stage familial osteoarthritis transcribes
 RT type II collagen mRNA encoding a cysteine in position 519.";
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
 RN [22]
 RP VARIANT SEDC ARG-285.
 RX MEDLINE=93252400; PubMed=8486375;
 RA Viikula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
 RA Peltonen L.;
 RT "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia.";
 RL Genomics 16:282-285(1993).
 RN [23]
 RP VARIANT SEDC CVS-206.
 RX MEDLINE=94063862; PubMed=8244341;
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75-->Cys mutation in the procollagen type II gene
 RT (COL2A1).";
 RL Hum. Genet. 92:499-505(1993).
 RN [24]
 RP VARIANT SEDC CVS-920.
 RX MEDLINE=93315508; PubMed=8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.;
 RT "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
 RT dysplasia.";
 RL J. Biol. Chem. 268:15238-15245(1993).
 RN [25]
 RP VARIANT SEDC SER-1128.
 RX MEDLINE=93140139; PubMed=8423604;
 RA Cole W.G., Hall R.K., Rogers J.G.;
 RT "The clinical features of spondyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 997 by serine in the alpha
 RT 1(II) chain of type II collagen.";
 RL J. Med. Genet. 30:27-35(1993).
 RN [26]
 RP Alignment Scores:
 Pred. No.: 5.62e-07 Length: 1418
 Score: 305.00 Matches: 302
 Percent Similarity: 21.10% Conservative: 64
 Best Local Similarity: 25.66% Mismatches: 376
 Query Match: 3.27% Indels: 437
 DB: 1 Gaps: 68
 US-09-931-704-3 (1-5087) x CA12_HUMAN (1-1418)
 QY 4305 GGGGATAGGAGGACCAACACCCTCCAGAGAGTGGGCGAGCCAAAGGCTTACCCCG 4246
 Db 60 GlycylGlnGlyProArgGlyAspArgGlyAspGlyGlyGlyAlaProGlyPro 79

QY 2915 CAGCAAGGAGAAATGAGCCTTCAAAACACTGTCGGGCTTGGCAAGAGGTGCTGACGACC 2856
Db 539 -----ArgGlyGlyAlaGlyProPr 545
QY 2855 TTGAGAGAACCAATTGCTGGGCTTGGAGGAGCAGATTGCAAGGGGTTAAGAGTGAAGTGG 2796
Db 545 o-----GlyPro-----GluGlyGlyLysGlyAlaAlaG 555
QY 2795 GTGAGAGAGAGTGGAGCAGAGGTGTAGACCCTCTTGGAGAAATTTGGCAGGAAA 2736
Db 555 yProProGlyProProGlySerAlaGlyThrProGlyLeu-----GlnGlyLe 571
QY 2735 GCCAGGAGAGAAA-----AAGGTCGGGAAGT 2709
Db 571 tProGlyGluArgGlyGlyProGlyGlyProGlyProGlyAspLysGlyGlyProG 591
QY 2708 AGAGAGGCGTGTGAGTGAACAAAGGTTTCTGCACAAAGAAACAGAGAGATGAGAG 2649
Db 591 ySerSerGlyValAspGlyAlaProGly-----LysAspGlyProArgGlyProTh 608
QY 2648 TGGCCCATTTGGAACAGAGCCCCATTTGCCGCTGACACACAGCCAGCCACAGCTGAGGG 2589
Db 608 rGlyProIleGlyProProGly-----ProAlaGlyGlnProGlyAspLysGlyGlyLe 626
QY 2588 GCTGGGAGAGACGGGTAC---TCACTAACACTCCGAAGAGAGGTTTCAGCTGATAGGAG 2532
Db 626 x--GlyAlaProGlyValProGlyIleAlaGlyProArgGlyGlyProGlyGluArg 645
QY 2531 CAGGGTT-----GCAGAGGAAGAAATGAGGAGC 2502
Db 645 yGluGlnGlyProProGlyProAlaGlyPheProGlyAlaProGlyGlnAsnGly----- 663
QY 2501 TCTCATGAGTAGGACAGAGAGAGGCTATGAACAGACAGCATGTTATAGGAAGAGGGCAG 2442
Db 664 -----GluProGly-AlaLysGlyGluArgGlyAlaProGlyGluLysGlyG 679
QY 2441 AGCCCTTTCCAAACCTCGGGGAGCGCTGGGCCCGGAGGAAAGGATGAGGAGATGTA 2382
Db 679 lu-----GlyGlyProProGlyAlaAlaGlyProAlaGlyGlySerGlyP 694
QY 2381 TAACTGGGAAGAAGGAG-----AAGCAAGGGGAGAGG-----CTG 2344
Db 694 roAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySerProGlyGlyProG 714
QY 2343 GGATA-----GGCAAGGGCCAGGGGAGGTTTCACACGCTGACAAGCACA 2299
Db 714 lAlaAlaGlyPheProGlyGlyArgGlyProProGlyPro----- 727
QY 2298 GAGACAACTAATTTCTCGGTATCTCGGGGTGGAGCCAGCTCCCGCCACCTTCT 2239
Db 728 -----ProGlySerAsnGlyAsn-ProGlyProProGly 738
QY 2238 TCCCTTTTCTTCTCCCTCCAGAGTAGGCTCTGGGATCTGGAGAGGAAACAGGGAAG 2179
Db 739 SerSerGlyAlaPro-----GlyLys 745
QY 2178 AGGGGCGCGGT---CCCCATGTTGCTCGGCTCTCCACCCCTCTCTCTCTCTG 2122
Db 746 AspGlyProProGlyProProGlySerAsnGlyAlaProGly---SerProGlyLysSer 764
QY 2121 GGTCTCTGACACCTCTGACACACCCCAATCTTACAGCATAACTCCATCCACAC 2062
Db 765 GlyProLysGlyAspSerGlyPro----- 772
QY 2061 CTAACCCCTCGCCACTACTGCTGGGTTCCCTACAGTCCCTTCCCTCCATGACCCCT 2002
Db 773 -----ProGlyGluArgGlyAlaProGlyProGlnGlyProPro 785
QY 2001 AGGAGCCACAGTAGATGGAGAGGAGCAGAGGAGGCC----- 1959
Db 786 -GlyAlaProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyPr 805

QY 1958 -----CCTGGGTCCAGAGCAGCA---GGGCCCCAG---GTCTGGGGGAGAGCGC 1915
Db 805 oProGlyMetProGlyAlaArgGlySerProGlyProGlnGlyLysGlyGlyGluAsn-- 824
QY 1914 AGCTGAACCCAGGCACTGGGCGGCCCTCTGTCTCCCAACCCCGCCCGCGCA 1855
Db 825 -----GlyLysProGlyPro----- 829
QY 1854 AGAGACACCAACACAGCGTGTAAACAGAGAGACTTCCCGGCCCCAGGGGAGGCGGG 1795
Db 830 ---SerGlyGlnAsnGly---GluArgGlyPro---ProGlyProGlnGlyLeuProG 846
QY 1794 C-----GCTGGGCTGGCAGCCAGTGGGGCAATGGCCCAAGGGGTGGCAGGCTGCC 1741
Db 846 yLeuAlaGlyThrAlaGlyGlu-ProGlyArgAspGlyAsnProGlySerAspGlyLeuP 866
QY 1740 CATATCCCTAGTCTCTGGGCCAGGAGTGGGGGCGCCACCTTAAGAGAGGAAGGG 1681
Db 866 ro-----GlyArgAspGlyAla-----ProGlyAlaL 875
QY 1680 CAGGAGATAGACAGACAGGGGAGCTGTGGAGCCCAAGAGCCCTCTCTCACCCCATCA 1621
Db 875 ysGlyAspArgGlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisPro--- 893
QY 1620 TTTACCAAGAACCTCCA-----CGGTTAGGACTGTCTTCCCAACTCTCTCAC 1573
Db 894 -----GlyProProGlyProProAlaGlyPheProGlyLysSerGlyAspArgGlyLut 911
QY 1572 TCCTCCCAACTCCAGATTCCTACCTGATCTACATAGTCTCCAGGAGGCGCCAGGCTCG 1513
Db 911 hrGlyPro-----AlaGlyProSerGlyAlaPro- 920
QY 1512 AGTTGGTCTCAGGTAGCGGTGAGGTCTAGTTTCTGATGAGGGGCGCCAGGCTCT 1453
Db 921 -----GlyProAlaGlySer-----ArgGlyProProGlyPro 931
QY 1452 GGGTCCCTGTGCGATTGAGAGCTGGCACTGCAGGAGGTGCGCAGAGCACCGTGACAGG 1393
Db 932 GlnGlyProArgGlyAspLysGlyGluThrGlyGluArgGlyAlaMetGlyLysGly 951
QY 1392 CACGCTAACATCCCC-----CACGAG 1372
Db 952 HisArgGlyPheProGlyAsnProGlyAlaProGlySerProGlyProAlaGlyHisGln 971
QY 1371 TCCCTGTGGGAGGATGGACGAGAGCATGAGCGGCGCCG-----GGCCCCA 1324
Db 972 GlyAlaValGlySerProGlyProAlaGlyProArgGlyProValGlyProSerGlyPro 991
QY 1323 CACGGGAA-----GCAGCTGGCAGCCCTCAGGTGTCTGTCTCCACCTTGGGCTGGGT 1270
Db 992 ProGlyLysAspGlyAlaSerGlyHisProGlyProIle--GlyProPro----- 1007
QY 1269 GGGGGTCAAGGTGGGACACGATGGGAGGAGGAGGATGGAGGGGAA 1217
Db 1008 --GlyProArg-----GlyAsnArgGlyGluArgGlySerGlu 1019
RESULT 15
DRPL RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN DRPLA
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;

RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
 RA Ashworth R.G., Ross C.A.:
 RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
 RT homologue";
 RL Neurobiol. Dis. 2:129-138 (1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
 RX MEDLINE=96081227; PubMed=8541849;
 RA Schmitt I., Epplen J.T., Riess O.;
 RT "Predominant neuronal expression of the gene responsible for
 RT dentatorubral-pallidoluysian atrophy (DRPLA) in rat";
 RL Hum. Mol. Genet. 4:1619-1624 (1995).
 CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
 CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
 CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
 CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U31777; AAA80337.1; --
 DR EMBL; X89453; CAA61623.1; --
 DR InterPro; IPR002951; Atrophin.
 DR Pfam; PF03154; Atrophin-1; 2.
 DR PRINTS; PR01222; ATROPHIN.
 FT DOMAIN 165 171 POLY-PRO.
 FT DOMAIN 303 306 POLY-PRO.
 FT DOMAIN 377 383 POLY-SER.
 FT DOMAIN 387 391 POLY-SER.
 FT DOMAIN 440 446 POLY-PRO.
 FT DOMAIN 477 480 POLY-HIS.
 FT DOMAIN 481 489 POLY-GLN.
 FT DOMAIN 502 505 POLY-PRO.
 FT DOMAIN 562 572 POLY-SER.
 FT DOMAIN 702 705 POLY-PRO.
 FT CONFLICT 455 455 N -> S (IN REF. 2).
 FT CONFLICT 594 594 F -> L (IN REF. 2).
 FT CONFLICT 689 689 P -> R (IN REF. 2).
 FT CONFLICT 717 717 T -> M (IN REF. 2).
 FT CONFLICT 737 737 A -> V (IN REF. 2).
 FT CONFLICT 965 965 MISSING (IN REF. 2).
 SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Alignment Scores:

Pred. No.:	6e-07	Length:	1183
Score:	304.50	Matches:	327
Percent Similarity:	28.73%	Conservative:	95
Best Local Similarity:	22.26%	Mismatches:	437
Query Match:	3.23%	Indels:	610
DB:	1	Gaps:	71

US-09-931-704-3 (1-5087) x DRPLA_RAT (1-1183)

Qy	1090	CAAAGACTCTCTCAGTCGACACAG-----ATCATCTCAAGTATCT	1134
Db	120	ProArgAspIleAspGlnAspAsnArgSerThrSerProSerIleTySerProGlySer	139
Qy	1135	ACTGATAGACATCTCGTCGTCGAAGTCCTCATCTCAACATACACATCGCTCTCTTC	1194
Db	140	ValGluAsnAspSer-----AspSerSerGlyLeuSerGln	152
Qy	1195	TCCTCCGCTTCCAGGAGTGTTCCTCCCTCCATCCCTCGCTCCATCTCGTGTGC	1254
Db	153	GlyProAlaArgProTyHisProProProLeuPhePro-----ProSerProPro	169
Qy	1255	CCACCC-----TCACCCCAACCCAGCCAGGTGGGACAGACACCTGAGGGGCTGCCAGC	1311

Db	170	ProProAspSerIleProArgGlnProGluSerGlyPheGluPro-----	184
Qy	1312	TGCTTCCCTGTTGGGCGCGCGCTCATGCTTCTCGTCCATCTCTGCCACAGGGA	1371
Db	185	-----HisPro-----	186
Qy	1372	CTCGTGGGATGTTAGCTGCCT-----GTGCAGGTGCTCTGGGACCTCCCTGC	1422
Db	187	-----SerValProThrGlyTyHisAlaProMetGluProProThr	201
Qy	1423	AGTCCAGCTCTCAATCGCACAGGGGCCAGGGCCTGCCCTCCATCCAGAAACCTA	1482
Db	202	SerArgLeuPheGlnGly-----	207
Qy	1483	TGACCTCACCGCTACCTGGAGCACCACTCCGAGCTTGGCTGGGACCTATGTAGTAT	1542
Db	208	-----ProProGlyAlaProPro-----His	216
Qy	1543	CCAGCTAGGAATCTGGAGTTGGGAGGAGTAGGAGTTGGGAAAGACAGTCTTACC	1602
Db	217	ProGlnLeuTyProGlySerAlaGlyGly-----GlyValleu-----	229
Qy	1603	GTGGAGGTTCTGTAATGATGGGTGAGGAGGG-----	1638
Db	230	-----SerGlyProProMetGlyProGlyGlyAlaAlaAlaSerSerValGly	246
Qy	1639	---CTCTTTGGCTCCACCACTCCCTGTCTGTGTCTATCTCTGCCCTTCCCTTAGG	1695
Db	247	ProProSerGlyGlyLysGlnHisPro-----	255
Qy	1696	TGGCCCCCACTTCCCATCCCTGGGCCAGGACTAGGATGGGAGGCTCGCACC	1755
Db	256	---ProProThrThrProIleSerSerGlyAlaSerGlyAlaPro-----	272
Qy	1756	CGCCTTGGCCCATTTGCCCACTGGCTGCAGCCAGCCCGCTCCCTCGGGGCG	1815
Db	273	-----ProAlaLysProProAsnThrProValGlyAla	283
Qy	1816	GGGAAGTCTCCT-----CTGTTTACCCGTGTGTGTGTCTCTTC	1857
Db	284	GlyAsnLeuProSerAlaProProAlaThrPhePro-HisValThrProAsnLeu--	302
Qy	1858	CGGGCGGGTGGTGGGACAGGGGCCACCTCCATGCTCGTGGTTCAG-----	1912
Db	303	-----ProProProAlaLeuArgProLeu	311
Qy	1913	---CTCGCCTTGCCTCCAGACCTGGGCGCTGTCTGTGACCCAGGGGCTCCCTT	1968
Db	311	uAsnAsnAlaSerAlaSerProProGly-----MetGlyAlaGlnProIleProGln	328
Qy	1969	CCGTCTGCTCTCCCATCTAGTGGCTCTAGGGGGTCTAGGGGAGAGGGAGTGT	2028
Db	328	yHisLeuProSerProHis-----AlaMetGlyGlnGlyMetSe	341
Qy	2029	AGGAAACCCAGGACAGTGGGCGGGTTAGGTGGATGGAGGTATCTGTAAAG	2088
Db	341	rglyLeuPro-----	344
Qy	2089	ATTTGGGTGTCTCCAGAGTGTTCAGAGAGCCAGGAGAGGAAGAGGGTGGAGG	2148
Db	345	-----Pro-GlyProGlyLysGlyProThrLeuA	354
Qy	2149	AGCCGAGGACACATGGGAAACCGGCCCTCTTCCTGCTTCACATCCAGAG--	2207
Db	354	LaProSerProHisProLeuProAlaSerSerAlaProGlyProProMetArgT	374
Qy	2208	---CCCTACTCTCGACCCAGGNA-----AGAAAGGAGAGAGGTGGCG	2250
Db	374	yrProTySerSerCysSerSerSerValAlaAlaSerSerSerSerAlaAlaT	394
Qy	2251	GGGAGTGGCTCCAGCC-----CCAGGATACACCGAGGAAATAGTTTCTCTC	2298
Db	394	hrSerGlnTyProAlaSerGlnThrLeuProSerTyProHisSerPhe-----	410

[illegible]

QY	3349	CCTCTCTTTT	CACAGCTGAACTACTACCTGGGCCCCCTTCAACAGCAGCCAGACTTCAACCC	3408
DB	623	-----	-----ThrAlaSerProProGlyProPro-----GlnTyrSerLysArgAlaPro	637
QY	3409	TCC	-----CCGCCTGGGGCAGAGACTCTGCCCAGGGCCAC	3444
DB	638	SerProGlySerTyrLysThrAlaThrProProGlyTyrLysProGly	-----	653
QY	3445	TGTTGACTTGGAGGTGGCGAAGCCTCAATGACAAACTGCGGCTGACCCAGAACTACGA	3504	
DB	654	-----	-----SerProSerPheArgThrGlyThrProProGlyTyrArg	667
QY	3505	GGCTACAGCCACCTTCTGTGTACTTGGTGGCTCAACCGTCAGCTGCCACTGCTGA	3564	
DB	668	GlyThrSerProPro	-----AlaGlyPro-----GlyThrPheLysPro	680
QY	3565	GCTGGCGCCAG	-----CCTGGCCCACTTCTGCACACGCTTCAGGCGCTGCTGGGCAG	3618
DB	681	GlySerProThrValGlyProGlyProLeu	-----ProProAlaGlyProSer	696
QY	3619	CATTGGGGCGTCATGGCAGCTCTGGGCTACCACTGCCAGCCCGCTGCCTGGGACTGA	3678	
DB	697	-----	-----SerLeuSerSerLeuProProProAlaAla	707
QY	3679	ACCACCTTGGACTCTCTGGCCCTGCCACAGTACTCTCTCCAGAAGATGGACGACTTCTG	3738	
DB	708	-----	-----ProThrThrGlyProProLeu	714
QY	3739	GCTGCTGAAGAGCTGCAGACCTGGCTGTGGCGCTGGCCAAAGGACTTCAACCGGCTCAA	3798	
DB	715	ThrAlaThrGlnIleLysGln	-----GluProAlaGlu	725
QY	3799	GAAGAAGATGCAGCCTCCAGCAGCTGCAGTCACCTCGACCTGGGGGCTCATGGCTTCTG	3858	
DB	726	GluTyrGluThrProGluSer	-----ProValProPro	736
QY	3859	ACTTCTGACCTTCTCTCTTCTGCTCCCTCCCTTCAAAACCTGCTCCCACTTTGTGAGAGCCA	3918	
DB	737	-----	-----AlaArgSerProProProLysValValAspVal	749
QY	3919	GCCTGTATGCCAACACCTGTTGACCGCAGGACAGAGCTGTGAGCCTCTGCCCTTTC	3978	
DB	750	ProSerHisAlaSerGlnSer	-----AlaArgPheAsnLys-----His	762
QY	3979	CTGGACCGCTGGGGCGTGTGATGCGATCAGCCCTGTCTCTCCCACTCCCAAGGCTCT	4038	
DB	763	LeuAspArgGlyPheAsnSerCysAlaArgSerAspLeuTyrPheValProLeuGluGly	782	
QY	4039	ACCGAGCTGGGAGAGGTACATAGGCCCTGCTCTCTGCTCTGTCTACAGGAAGTCATG	4098	
DB	783	SerLysLeuAlaLysLysArgAlaAsp	-----LeuValGlu-LysVa	796
QY	4099	CTCAGAGGAGTGTGAA	-----	4114
DB	796	LargeArgGluAlaGluInArgAlaArgGluLulysGluArgGluArgGluArgGluArg	816	
QY	4115	-----	-----GTGGTTTCAGGTTGGTGC	4131
DB	816	gGluLysGluArgGluArgGluLysGluArgGluLeuGluArgSerValLysLeuAlaGlu	836	
QY	4132	AGAGCGCTCATGGCTCTCTGCTTCTTGCCTACCACTTGGCCAGTGGCCACCAGCCCTCT	4191	
DB	836	nGluGly	-----	838
QY	4192	CAGTGGCACATCTGGAGGGCAGGGTTTGAG	-----GGGCCACCCACACAT	4239
DB	839	-----	-----ArgAlaProValGluCysProSerLeuGlyProValProHisAr	853
QY	4240	GCCTTTCTGGGGTGAAGCCCTTTGCTGCTGCCCACTCTCTCTGGATGGGTGCTGCC	4297	
DB	853	gPro	-----ProPheGluProGlySerAlaValAlaThrVa	865
QY	4297	-----	-----	4297

Db 865 lProTyrLeuGlyProAspThrProAlaLeuArgThrLeuSerGluTyrAlaArgPr 885
Qy 4298 -----TTATCCCAATCACTTATACATCAATCAATCAAGAAACAAACATGGTGCAAT 4350
Db 885 ohisValMetSerProGlyAsnArgAsnHisPro-Phe----- 897
Qy 4351 TCTACAAAAAGAGATGATTAAACAGTGCAGGGTTGGGTCTGCATTGGAGTGCCCT 4410
Db 898 --TyrValProLeuGlyAlaValAspProGlyLeuLeuGlyTyrAsnValProAlaLeuT 917
Qy 4411 ATAAACACAGAGAGAAATACTGAAAGCAGGGGAGGAGCAGACACAGACAGACCCAG 4470
Db 917 yrSerSerAspProAlaAlaArgGluArgGluArgGluAlaArgGluArgAspLeuArgA 937
Qy 4471 GAGTCTCAAGACACAGAGTGGCAAAACCCGAGCTGAGCATCAGGACCTTGCTCG 4530
Db 937 spArgLeuLys-----ProGlyPheG 944
Qy 4531 AATTGCTTCCAGTATTACGGTGCCTCTTCTCTGCCCTTCCAGGGTATCTGTGGT 4590
Db 944 luValLysProSerGluLeuGluProLeu-----HisGlyV 956
Qy 4591 TGCCAGGCTGGGAGGGCAACCATAGCCACACACAGGA----- 4629
Db 956 alProGlyProGlyLeuAspProPheProArgHis-GlyGlyLeuAlaLeuGlnProGly 975
Qy 4630 -----TTCTCTGAAAGTTTACATGCAGTAGCATTTGGGGTGTAGG 4671
Db 976 ProProGlyLeuHisProPheProPheHisProSerLeuGlyProLeuGluArgGluArg 995
Qy 4672 GTGGCAGCTCCCAAGGCGCTCCGCCAGCCGCCACCCACTCATGACTCTAAGTGTGTTG 4731
Db 996 LeuAlaLeuAlaAlaGlyProAlaLeuArgPro----- 1006
Qy 4732 TATTAATATTATTATTGGAGATGTTATTATTATGATGATATTATTGTC---AGAATT 4788
Db 1007 -----AspMetSerTyrAlaGluArgLeu 1014
Qy 4789 TCT-----ATTCTGTATTAAACAAATAAATGCTTGCCCA 4824
Db 1015 AlaAlaGluArgGlnHisAlaGluArgValAlaAlaLeuGlyAsnAspProLeuAlaArg 1034
Qy 4825 GAACCTTAGTCTCTTTGCCAGCCTCACCCCTCCTGGTGTCTCATCAGACTCTTGCCACCCC 4884
Db 1035 LeuGlnMetLeuAsnValThrProHisHisHisGlnHisSerHisIleHisSerHisLeu 1054
Qy 4885 TGGCTCCCACTCCCTGCTTGCCTCTGGTGGAGCTGCACAGAGCTCTGGGAAGAGGCCCTC 4944
Db 1055 HisLeu-HisGlnGlnAspAlaIleHisAlaAla-----SerAlaSerValHisProLe 1072
Qy 4945 TTCCTCCCGCACTGGGGCGATGGGCGCACCTCAGACTTACCCACTGCTGCTGCCACC-- 5002
Db 1072 ulleAspProLeuAlaSerGlySerHisLeuThrArgIleProTyrProAlaGlyThrLe 1092
Qy 5003 -ACCAACCCCTTGATCCCTCAGTCCCTCCACACAGCTTCTGTG------CACCCAGGTT 5055
Db 1092 uProAsnProLeuProHisProLeuHisGluAsnGluValLeuArgHisGlnLeuPh 1112
Qy 5056 TCCCTCACCCAC 5068
Db 1112 eAlaAlaProTyr 1116

Search completed: January 27, 2003, 16:05:35
Job time : 244.026 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:48:57 ; Search time 318.364 Seconds
(without alignments)
6584.667 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aactgcgagtgaggcctggc.....ccttgctaagtctctcca 5087

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09931704/runat_27012003_154126_3635/app_query.fasta_1.7189
-DB=SPTREMBL_21 -QMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPWT=spt -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704@cgn_1.1.633 @runat_27012003_154126_3635 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mbc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	896	9.5	225	4 Q9UBD9	Q9ubd9 homo sapien

ID	Q9UBD9	PRELIMINARY;	PRT;	225 AA.
AC	Q9UBD9;			
DT	01-MAY-2000 (TRENBLrel. 13, Created)			
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)			
DE	Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like cytokine) (Similar to cardiotrophin-like cytokine, neurotrophin-1/B-cell stimulating factor-3).			
DE	CLC.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9432254; PubMed=10500198;			
RA	Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lille J., Scully S., Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F., Simonet W.S., Boone T., Chang M.-S.			
RA	"Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the IL-6 family."			
RT	Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).			
RL	[2]			

ALIGNMENTS

RESULT 1

Q9UBD9 ID Q9UBD9 PRELIMINARY; PRT; 225 AA.

AC Q9UBD9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like cytokine) (Similar to cardiotrophin-like cytokine, neurotrophin-1/B-cell stimulating factor-3).

2	863	9.1	225	11	Q9QZM3
3	420	4.5	5146	6	Q8SPM4
4	404	4.3	4123	4	O75851
5	397.5	4.2	1315	10	Q9SPM0
6	396.5	4.2	763	2	Q9XDH2
7	382.5	4.1	1047	5	Q9VR13
8	378.5	4.0	2284	5	Q9VPG1
9	374.5	4.0	1480	10	Q9LIE8
10	373	4.0	676	6	Q95JC9
11	370.5	3.9	1548	4	Q9NYW9
12	370.5	3.9	2161	4	Q9Y566
13	369	4.0	1458	13	Q910B9
14	358.5	3.8	1616	4	O15054
15	354	3.8	2158	11	Q9WU13
16	351.5	3.7	2087	11	Q9WU88
17	351	3.7	2167	11	Q9WV48
18	341.5	3.6	2703	5	Q9VEG7
19	339.5	3.6	1491	13	Q91718
20	336.5	3.6	1003	4	Q96DN6
21	336	3.6	1188	10	Q41805
22	336	3.6	2715	5	O61603
23	333.5	3.5	967	4	Q96Q00
24	333	3.6	774	12	O41971
25	333	3.6	1414	5	Q26634
26	333	3.5	3084	12	Q8UZ11
27	331.5	3.6	1486	13	Q91717
28	329.5	3.5	1289	10	Q9FLQ7
29	329.5	3.5	3570	4	Q99552
30	327.5	3.5	706	12	O41972
31	327	3.5	1418	13	Q9W7R9
32	326.5	3.5	2157	4	Q95875
33	326	3.5	1953	5	Q9BIT7
34	325.5	3.5	1182	4	Q99495
35	325	3.5	1449	13	Q910C0
36	323	3.4	2187	11	P70670
37	321.5	3.4	1608	4	Q96RKO
38	320	3.4	2157	4	Q96QC6
39	319.5	3.4	1006	10	Q9LMQ1
40	319	3.4	1453	4	Q9Y6T1
41	318.5	3.4	2971	4	Q9Y5L9
42	318	3.4	1378	5	Q97405
43	315	3.3	566	6	Q95JDI
44	314.5	3.4	1447	13	Q9IB91
45	314	3.4	1487	4	Q14047

```

RX MEDLINE=9382254; PubMed=10448081;
RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,
RA Ruben S., Alderson R.F.;
RT "Computational EST database analysis identifies a novel member of the
RT neuroplectic cytokine family.";
RL Biochem. Biophys. Res. Commun. 262:132-138(1999).
RN (3)
RP SEQUENCE FROM N.A.
RA Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176912; AAF00992.1; -
DR EMBL; AF176912; AAD54284.1; -
DR EMBL; AF176911; AAF00991.1; -
DR EMBL; AY049779; AAL15436.1; -
DR EMBL; BC012933; AAL12939.1; -
SQ SEQUENCE 225 AA; 251176 MW; E2DD4B6280833B55 CRC64;

Alignment Scores:
Pred. No.: 1-96e-55 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 4 Gaps: 2

US-09-931-704-3 (1-5087) x Q9UBD9 (1-225)
QY 3311 GGTGAGCGAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTACAGCTGAACCTACCTGGGCGGCGGCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTGGGCGAGAGACTGTGCCAGGCGACCTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGGCGAGCCTCAATGACAACTGGGCTGACCCAGAACTACGAGGCGCTACAGC 3514
Db 92 GluValTyrArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGCGCCCACTTCTGCACAGCCTCCAGGCGCTGTGGCAGCATTTGGCGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCACTGCCCGCGCTGCTGGAGCTGAACCACTTGGACTCCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTyrPro 171
QY 3695 GGCCTCTCCACAGTACTTCTCCAGAAGATGGACGACTTCTGGCTGCTGAAGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTyrLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGTGGCGCTGGCCAAAGACTTCAACCGGCTCAAGAGAGATGCAGCT 3814
Db 192 GlnThrTyrLeuTyrArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGCAGCTGAGTCAACCTGACCTGGGCGCTCATGGCTTC 3856
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

```

RESULT 2

```

Q9QZM3 PRELIMINARY; PRT; 225 AA.
AC Q9QZM3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
DE Neurotrophin-1/B-cell stimulating factor-3.
GN BSF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99432254; PubMed=10500198;
RA Senaldi G., Varnum B.C., Sarmiento U., Starnes C., Lile J., Scully S.,
RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
RA Simonet W.S., Boone T., Chang M.-S.;
RT "Novel neurotrophin-1/B cell-stimulating factor-3: a cytokine of the
RT IL-6 family.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).
DR EMBL; AF176913; AAR00993.1; -
DR MGD; MGI:1930086; Bsf3.
SQ SEQUENCE 225 AA; 25261 MW; 68B1FEAAB7F1A950 CRC64;

Alignment Scores:
Pred. No.: 4-38e-53 Length: 225
Score: 863.00 Matches: 163
Percent Similarity: 88.66% Conservative: 9
Best Local Similarity: 84.02% Mismatches: 10
Query Match: 9.15% Indels: 12
DB: 11 Gaps: 2

US-09-931-704-3 (1-5087) x Q9QZM3 (1-225)
QY 3311 GGTGAGCGAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTACAGCTGAACCTACCTGGGCGGCGGCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTGGGCGAGAGACTGTGCCAGGCGCCTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGGCGAGCCTCAATGACAACTGGGCTGACCCAGAACTACGAGGCGCTACAGC 3514
Db 92 GluValTyrArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGCTCAACCGTCAAGCTGCGCTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGCGCCCACTTCTGCACAGCCTCCAGGCGCTGTGGCAGCATTTGGCGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCACTGCCCGCGCTGCTGGAGCTGAACCACTTGGACTCCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTyrPro 171
QY 3695 GGCCTCTCCACAGTACTTCTCCAGAAGATGGACGACTTCTGGCTGCTGAAGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTyrLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGTGGCGCTGGCCAAAGACTTCAACCGGCTCAAGAGAGATGCAGCT 3814
Db 192 GlnThrTyrLeuTyrArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGCAGCTGAGTCAACCTGACCTGGGCGCTCATGGCTTC 3856

```


Db 4630 heArgLeuArgTrpArgGluAlaGlyIleProProGlyGlyGlyCysArgGlyProTrpA 4650
QY 4405 -----TGCCCTTATAAACCAAGAGAAATACTGAA 4435
Db 4650 laGlnThrGluSerCysAsnMetGlyProCysPro----- 4661
QY 4436 AGCACAGGGCAGGGAC-----AGACCAGACAGACCCAGG 4471
Db 4662 -----GlyGluSerCysGluAlaGlnAspThrValProThrProAspCysAlaAsnG 4679
QY 4472 AGTCTCAAAGCACAGAGTGCACAAACCCAGCTGAG----- 4512
Db 4679 inCysProArgSerCysValAspLeuTrpAspArgValGluCysLeuGlnGlyProCysA 4699
QY 4513 -----CATCAGGACCTTGCCTCGAATTGCTTC 4540
Db 4699 rgProGlyCysArgCysProProGlyGlnLeuValGlnAsp-----GlyHisCysValP 4717
QY 4541 CAGTATTACGGTGCCTCTTCTGCCCTTCCAGGGTATCTGTGGGTTGCCAGGCTG 4600
Db 4717 roValSerSerCysArgCysGlyLeuProSerProAsnAlaSerTrpAlaLeuAlaProA 4737
QY 4601 GGCAG-----GGCAACCATAGCCACACAGGATTTCTGAAAGTT 4642
Db 4737 laGluValValArgLeuAspCysArgAsnCysThrCysValAsnGly-----SerL 4754
QY 4643 TACAATGCAGTAGC-----ATTTTGGG----- 4665
Db 4754 euAlaCysSerSerHisGluCysProThrLeuGlyProTrpSerAlaTrpSerAsnCysS 4774
QY 4666 -----TGTAG-GGTGGCAGCTCCCAAGG----- 4688
Db 4774 exAlaProCysGlyGlyThrThrLysArgHisArgSerCysLysGluGlyProGlyV 4794
QY 4689 -----GlyProCysGlnAlaGlnAspMetGluGlnGlnAspCysAsnLeuGlnProCysP 4814
QY 4696 CCCAGCCCCCACCCTCATGAC-----TCTAAGTGTTGTGTTATTAATATTTATTTATTTGG 4752
Db 4814 roGluCysProProGlyGlnValLeuSerAlaCysAlaVal----- 4827
QY 4753 AGATGTTATTATTAGATGATATTATTGCAGAAATTTCTATTCTGTATTAAACAATAAA 4812
Db 4827 ----- 4827
QY 4813 ATGCTTGCCCGCAGAACTTAGTCTCTT----- 4838
Db 4828 --SerCysProArgLeuCysSerHisLeuGlnProGlyThrProCysMetGlnGluProC 4847
QY 4839 -----TGCCCGACCTCACCCCTCTGCTGCTCATCAGACTCTTGCCAC 4881
Db 4847 ysGlnLeuGlyCysAspCysProArgGlyGlnLeuLeuHisAsnGlyThrCysValProp 4867
QY 4882 CCTGTGCTCCACTCCCTGCG-----TTGCTCTGG 4911
Db 4867 ro---AlaGluCysProCysThrGlnLeuSerLeuProTrpGlyLeuThrLeuThrLeuG 4886
QY 4912 TGGAGCTGCACAGAGCTCTGGGAAGAGGCGCTCTCTCTCCCGCAGCTGGGCGATGGCG 4971
Db 4886 luGluGlnHisArg-----GluLeuProPro-----GlyThrLeu-Leu 4898
QY 4972 CACCTCAGACTTACCACTGCTGCTGCCACCAACCCCTTGATCCCTCAGTCCCTCCCA 5031
Db 4899 ThrGlnAsnCysThrHisCysIleCysGlnGlyAlaPheSerCysSerLeuThrAsp 4918
QY 5032 CACAGCTTCTGTCACCC 5049
Db 4919 CysGlnGluCysProPro 4924
RESULT 4
O75851
ID O75851 PRELIMINARY; PRT; 4123 AA.

AC O75851;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE WUGSC:H DJ0751H13.1 protein (fragment).
GN WUGSC:H DJ0751H13.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
SEQUENCE FROM N.A.
RA Leonard S., Graves T., Strommatt C.;
RT "The sequence of Homo sapiens PAC clone RP4-751H13.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC004877; AAC36301.1; -.
DR HSSP; P01130; 1AJT.
DR InterPro; IPR000923; BlueCu_1.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000421; FAS8_C.
DR InterPro; IPR001092; HLH Basic.
DR InterPro; IPR002223; Kunitz BPTI.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00754; F5_P8_type_C; 1.
DR Pfam; PF00057; ldl_recept_a; 11.
DR Pfam; PF01826; TIL; 5.
DR Pfam; PF00090; tsp_1; 14.
DR Pfam; PF00094; vwd; 3.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00231; FAS8C; 1.
DR SMART; SM00192; LDLa; 10.
DR SMART; SM00209; TSPI; 14.
DR SMART; SM00214; VWC; 1.
DR SMART; SM00216; VWD; 3.
DR PROSITE; PS00280; BPTI_KUNITZ_1; UNKNOWN_1.
DR PROSITE; PS00196; COPPER_BLUE; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01209; LDLRA_1; 9.
DR PROSITE; PS00068; LDLRA_2; 18.
DR PROSITE; PS00092; TSPI; 22.
KW Glycoprotein.
FT NON TER
SQ SEQUENCE 4123 AA; 434981 MW; 7AAB6F8DCE012FB CRC64;

Alignment Scores:
Pred. No.: 4.11e-20 Length: 4123
Score: 404.00 Matches: 514
Percent Similarity: 27.68% Conservative: 142
Best Local Similarity: 21.69% Mismatches: 711
Query Match: 4.28% Indels: 1006
DB: 4 Gaps: 131

US-09-931-704-3 (1-5087) x O75851 (1-4123)

QY 18 GCGCGATGGGATTTATAAGCTTCGCCGAGCGCGGCTCGCCCT----- 62
Db 1309 GlyValTrpAspCysProAsp-----GlyAlaAspGluGlyProGlyHisCysProLeu 1326
QY 63 -----CCCACTCCCGCAGCTCCCGGAGAGCGCGACCGCGCGCGCGAGCCCC 113
Db 1327 ProSerLeuProThrProProAlaSer-----ThrLeuProGly---ProSerPro 1342
QY 114 AGCCCCATGGACCTCCCGAGCAGGTGTGAACCAACCAACTAGCCCTGCTCTTCATACATGA 173

Db 1343 GlySerLeuAsp----- 1346
QY 174 CAAGCAGGCCCATCTGATACCTAAACCGACCAAGTACACAGCCCTCCAACTCACTCTCT 233
Db 1347 -----ThrAlaSerSerProLeuAlaSerAlaSer 1356
QY 234 GCCTGCCCA-----GACCTCACACATCC 257
Db 1357 ProAlaProProCysGlyProPheGluPheArgCysGlySerGlyGluCysThrProArg 1376
QY 258 TTGTGG-----ACTCAACCTCAACCGCACTAAATCAACAAATCCCAAGCTAAACTA 311
Db 1377 GlyTirpCysAspGlnGluAspCysAlaAspGlySerAspGluArgGly----- 1394
QY 312 ATCTGAACCTTTAAAGTAACCCAGTCTTAAACCTAACCTAGCCCAATGCAATTATAT 371
Db 1395 -----CysGlyGlyProCysAlaProHisAlaProCysAla----- 1407
QY 372 CTACCTAGCCAAACCTTAAGTCTTTCAGTCCCAAGTG-----TCCACTGAATCC 425
Db 1408 -----ArgGlyProHisCysValSerProGluGlnLeuCysAspGlyValArgGln 1424
QY 426 TCACCTTGCTCTCACTCAAAATCCAGAAAGCATATTTCCCACTGCCACATCCCTC 485
Db 1425 CysProAspGlySerAspGluProAsp-----AlaCys 1436
QY 486 CTTACAGCACCACCTCGGCTCTGGACTCTGTGTATCTCTGGATGTCCTCAACTCTGCA 545
Db 1437 ValGluAlaProAlaPro-----ProAlaMeta*rglyProProGly----- 1450
QY 546 GTGCCATCAGCAACAAGCCGACTCGTCAATGACACTCTCTCTCTCTCTCTCTCTCTCT 605
Db 1451 -----GlnAlaGlyProThrSerSerArgAlaPro-SerProProSerProProG 1468
QY 606 CTTGAGGCTGATGAAAGGCTCATTAAGTCCAACTTTTCCCACTTAACACCAAGAA 665
Db 1468 uAlaGlnGlyGluGlyArgGlyGlnGluArgSerArgThr-HisLeuThrValProA 1488
QY 666 CGGGTGAACTCCACACTGCCACGTCCTCTGAGAGTGAGGACTTAATCTCTCTCAATC 725
Db 1488 laGly---SerThrGlnLeuPro----- 1494
QY 726 TAAACCCACCTTACACTTCCACACTCAGGAATCACATCTAGATAATACCAAACTAA 785
Db 1495 -----LeuCysProGlyLeuP 1500
QY 786 GCCCATAAGGC---AGCCCGACCTAGTGGTCTAAACCTATACCTTG----- 830
Db 1500 heProCys-GlyValAlaProGlyLeuCysLeuThrProGluGlnLeuCysAspGlyIle 1519
QY 831 -----CTTCCTATGGTGAGTCTGTTCTTGGGCGCGCTCTCTCTCTCTCTCTCTCT 884
Db 1520 ProAspCysProGlnGlyGluAspGluLeu----- 1529
QY 885 AGAGCTGACTGTGCTCAGCTGCCAGCTCTG----- 915
Db 1530 -----AspCysGly-GlyLeuProAlaLeuGlyGlyProAsnArgThrGlyLeuProC 1547
QY 916 -----ACATGTGTGTCTCCCAACCTCTGACTCCCTCCCTCAAGCTGCACTGGGAGTG 965
Db 1547 sProGluTyThrCys----- 1552
QY 966 GAAGACTGGCAGGAGCTAGGTACAACTGGAAACACAGGAGTCCAGCTGAGTCCCTA 1025
Db 1553 -ProAsnGlyThrCysIleGlyPheGlnLeu-ValCysAspGlyGlnProAspCysGlyA 1572
QY 1026 GGCCTGGCCCGCTCCCTCA-----TGTACACATATACATGCTTGG 1067
Db 1572 rgProGlyGlnValGlyProSerProGluGluGlnGlyCysGly-----A 1587
QY 1068 CACACACAGTGG-----CACATGCGCAAGACTCTCTCAGC 1106
Db 1587 laTrpGlyProTrpSerProTrpGlyProCysSerArgThrCysGlyProTrpGlyGlnG 1607
QY 1107 TGACACACAGATCCATTCTCAAGTATCTACTATAGACACTCATCGCTGCCAAGTCTCTCA 1166
Db 1607 lyArgSerArg-----ArgCysSerProLeuG 1616
QY 1167 TCCTCAACATACATACATGCTCTCTTCTCTCTCCGCTCTTGC----- 1207
Db 1616 lyLeu-----LeuValLeuGlnAsnCysProGlyProGluHisGln 1630
QY 1208 --GAGGAGTGTTCCT 1256
Db 1630 erGlnAlaCysPheThrAlaAlaCysProValAspGlyGluTrpSer-ThrTrpSerPro 1649
QY 1257 ---ACCCTCACCCCCAGCCCAAGGTGGGACACACACCTGAGGGGTGCGCAGCTG 1313
Db 1650 TrpSerValCysSerGluProCysArgGlyThrMetThrArgGlnArgGlnCysHis--- 1668
QY 1314 CTTCCCGTGTGGC-----CGGCGCGCGCTCATCTCTCTCTCTCTCTCTCTCTCTCT 1352
Db 1669 SerProGlnAsnGlyArgThrCysAlaAlaLeuProGlyGlyLeuHisSerThrArg 1688
QY 1353 CCATCC-----TGCCACAGGGGACTCGTGGGATCTTAGCGTCCGTGCACGGTGC 1406
Db 1689 GlnThrLysProCysProGln-----AspGlyCys-----ProAsnAlaThrCys 1703
QY 1407 TCTGGCACCTCCCTG-----CAGTGCCAGCTCTCAATCGCAGGGGACC 1451
Db 1704 SerGlyGluLeuMetPheGlnProCysAlaProCysProLeuThrCysAspAspIleSer 1723
QY 1452 CAGGCGCTGGCCCTCCATCCAGAAACCTATGACTCACCCTGCTACCTGGAGCAGCAAC 1511
Db 1724 GlyGlnValThrCysProAspTrpProCysGlySerProGlyCysTrp-CysProG 1743
QY 1512 TCCGAGCTT-----GGCTGGACCTATGTCAGTATCCAGCTCAGGAGTCTGG 1559
Db 1743 uGlyGlnValLeuGlySerGlyGlyTrp-----CysValTrpProArgGlnCysProC 1761
QY 1560 GAGTTGGGAGAGTGAGGAGTTGGGAAAGACAGCTCTCAACCGTGGAGGTT----- 1612
Db 1761 sLeuValAspGlyAlaArgTyTrp-ProGlyGln-----ArgIleLysAlaA 1777
QY 1613 -----CTGTAATATGAGGGTGAGAGG----- 1636
Db 1777 spCysGlnLeuCysIleCysGlnAspGlyArgProArgArgCysArgLeuAsnProAspC 1797
QY 1637 -----GGCTCTTGGCTCCACAGTCCCTCTGTGTCTATCTCTCTCTCTCTCTCTCT 1691
Db 1797 ysAlaGlyGluAlaLeuProSerGlySerLeuValLeuSerLeuAspArgProAlaA 1817
QY 1692 TAGGTGGCCCCCCT 1748
Db 1817 isProProProSerGlySerAspCysTrpProSerLeuSerGlyLeuValLeuVal 1837
QY 1749 TCGCACCCGCTTGGC-----CCATTGCCCACTGGCTGCCAGCCAGCCG----- 1794
Db 1837 euLeuValThrLeuGlyGlnValProGlyProLeuTrpLysProGluHisProVal 1857
QY 1795 -----CCGCGCT 1816
Db 1857 euProGluLeuGlnProProProLeuArgProArgSerProValProTrpHisProP 1877
QY 1817 GGAAGTCTCTCTGTTTACACCGTGTGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1863
Db 1877 roGlnGlyThrGlnThrGluProCysGlyGlyCysGluHisGlnGlyGlnValHisArg 1897
QY 1864 --GGGTTGGGTTGGGACAGAGGGGCCCC-----A 1891
Db 1897 alGlyGluArgTrp--HisGlyGlyProCysArgValCysGlnCysLeuHisAsnLeu 1916
QY 1892 CTCCCATCT 1935
Db 1916 hrAlaHisSerProTyThrCysProLeuGlySerCysProGlnGlyTrpValLeuValG 1936

Qy	1935	----	1935
Db	1936	luGlyThrGlyGluSerCysCysHisCysAlaLeuProGlyGluAsnGlnThrValGlnP	1956
Qy	1936	-----GGCCCTGCTGCTCTG-----GACCACGGGGCTCCCTCCGCTGCCT	1978
Db	1956	roMetAlaThrProAlaAlaAlaProAlaProSerProGlnIleAeRgPheProLeuAlaT	1976
Qy	1979	CTCCATCTAGCTGGCCCTCTCTAGGGGGTTCATGGGGAAAGGAGACTGTAGGGA	2033
Db	1976	hrTyrlleLeu-----ProProSerGly-----GlySerCysArgProLeuS	1990
Qy	2034	-----ACCAGGCAGTAGTGGCAGGGGGTTAGGTGTGTGATGAGGTTA	2078
Db	1990	erSerProThrProAla-----CysLeuSerLeuLeuHisProA	2003
Qy	2079	-----TGCTGTAAAGATTGGGGTGGTCCAGAGGTGTTCCAGAGAGCCCCAGGAGAGG	2133
Db	2003	spProCysTyrSerProLeuGlyLeuAlaGlyLeuAlaGluGlySerLeuHisAlaSerS	2023
Qy	2134	AAGAGGGTTGAGGAGCCGAGCCACCATGGGAAACCGGCCCTCTCCGCTGTTCCTC	2193
Db	2023	erGlnGlnLeuGluHisProThrGlnAlaLeuLeuGlyAlaProThrGlnGlyProS	2043
Qy	2194	TTCCA-----CATCCAGACCC-----T	2211
Db	2043	erProGlnGlyTrpHisAlaGlyGlyAspAlaTyralalysTrpHisThrArgProHist	2063
Qy	2212	ACTCTGGAGCCAGGGAAGAAAGGAAGAGGTGGCGGGAGCTGGCTCCAGGCCCCAG	2271
Db	2063	yrLeuGlnLeuAspLeuLeuGlnProArgAsnLeuThrGlylleLeuValProGluThrG	2083
Qy	2272	GA-----TACACCGAGGAAATTAGTTGTCTGTCTGTGTGCAGCGTGT-----	2315
Db	2083	lySerSerAsnAlaTyralaSerSerPheSerLeuGlnPheSerSerAsnGlyLeuHisT	2103
Qy	2316	-----GAACCTCCCGCTGGCGCTTGCCCTATCCAGCGCTCTCCCTTGCT	2361
Db	2103	rpHisAspTyraArgAspLeuLeuProGlylleLeuProleuProlysValSerProAla	2122
Qy	2362	-----TCTCCCTCTTTCCAGTTATACATCTCCCTCATCTCCCTTCC	2403
Db	2123	GlnGlyArgTrpGlyGlnGlnProThrMetProPheCysGlyPheHisSerLeuCysPro	2142
Qy	2404	CTGGCCCCAGCGCTCCCCGAGGTT-----	2430
Db	2143	GlnGlyProSerSerValProGluGlyHisGlyLeuHisSerMetLeuValGluTyLeu	2162
Qy	2431	-----TGAAAGGGCTCGCCCTCTCCCTATACCATG-----CTGTCT	2469
Db	2163	LeuPheProArgAsnTrpAspAspLeuAspProAlaValTrpThrPheGlyArgMetVal	2182
Qy	2470	TCCATAGCTTCTCTCTCTACTCATGAGACTGCTCCATTT-----CTTCTCTGCA	2525
Db	2183	GlnAlaArgPheValArgValTrpProHisAspValHisHisSerAspValProLeuGln	2202
Qy	2526	ACCTGTCTCTATCAGCTGAACCC-----TTCTTTCCGAGGTGTAGTAGTACCGCTCT	2579
Db	2203	ValGluLeuLeuGlyCysGluProGlyValGlyLeuArgCysAlaSerGlyGluCysVal	2222
Qy	2580	CTCCCCAGCCCTCAGCTGTGGGCT-----	2606
Db	2223	Leu-----ArgGlyGlyProCysAspGlyValLeuAspCysGluAspGlySer	2238
Qy	2607	-----GGTGT-----GTCAGCGCAAA	2624
Db	2239	AspGluGluGlyCysValLeuLeuProGluGlyThrGlyArgTyThrValAlaGlyArg	2258
Qy	2624	-----	2624
Db	2259	AlaAlaHisAlaLeuGlyLeuAlaPheGluGlyThrAlaMetTrpGluGlyProGlyThr	2278

[illegible]

Db 2620 -----ProCysSerGlyGlyThrGlyLysGlyValLeuGlyTrpGlyHisGlyGly 2636
Qy 3261 TGTGAGGGACACAAAATAGAACTTTGGGAGCAGCTATCTCTTGGTGAGCCAG 3320
Db 2637 -----SerThrValGlyThrGlyArgLeuGlyLeuProAlaPro 2649
Qy 3321 CGGCTC -----TGCCCTCTCTCT ----- 3338
Db 2650 ArgLeuThrTrpCysProSerProThrArgLeuArgAlaGlyProCysValCysGluCys 2669
Qy 3339 -----TCCCATCACC -----TCTCTTTT 3359
Db 2670 ArgSerValProGluGlyAlaGlyAlaProMetProThrLeuLeuProGlySerGlnGly 2689
Qy 3360 CACAGCTGAACCTACCTGGCC -----CCCTTTTCA 3389
Db 2690 GlnGlnLysLeuGlnTrpAlaLeuCysGlySerLeuProSerLeuLeuCysProLeuGly 2709
Qy 3390 ACGAGCCAGACT---TCAACCTCCCGCTGGGGGAGAGACTCTGCCCGGCGCACTG 3446
Db 2710 LeuSerAlaLeuPheHisLeuPro-----GlyCysArgCysProProGlyLeuLeu 2727
Qy 3447 TTG---ACTGGAGGTGGGAGCGCTCAATGACAAACTGGCGCTGACCCAGAACTAGG 3503
Db 2728 LeuHisAspThrArgCysLeuProLeuSerGluCysProCysLeuValGlyGluLeu 2747
Qy 3504 AGG---CCTACAGCCACCTTC-----TGTTGTTACTTGGCTGGCTCAACCGTC 3548
Db 2748 LysTrpProGlyValSerPheLeuLeuGlyAsnCysSerGlnCysValCysGluLysGly 2767
Qy 3549 AGGCTG-----CCACTGTCTGAGCTGGCGCGAGCTCTGGCGCCACT 3587
Db 2768 GluLeuLeuCysGlnProGlyCysProLeuProCysGlyTrpSerAlaTrpSerSer 2787
Qy 3588 TCTGACACGCTCCAGGCGCTGGGCGCATTTGGGCGCTCATGGCAGCTCTGGCT 3647
Db 2788 TrpAlaProCysAspArgSerCysGlySerGlyValArgAlaArgPheArg-Ser- 2805
Qy 3648 ACCCACTGCCCGCTGCTGGAGCTGACCCACTGGACTCTGGCGCTGCCACCA 3707
Db 2806 -ProSerAsnProAlaAlaTrp-----GlyGlyAlaProCys- 2818
Qy 3708 GTGACTTCTCCAGAGATGGACGACTTCTGGCTGCTGAAGCA-----GC 3752
Db 2819 -----GluGlyAspArgGlnGlnLeuGlnGlyCysHisThrValCysGly 2833
Qy 3753 TGCAGACTGGCTGGCGCTCGGCAAGCACTTCAACCGGCTCAAGAGAA- 3804
Db 2833 yThrGlyIleAlaGlySerLeuGlyAlaGlyValProProSerSerSerGlnPheCysTh 2853
Qy 3804 ----- 3804
Db 2853 rLeuArgThrHisGlyMetGlyProThrAspHisSerThrTrpGlyIleGluValPheGly 2873
Qy 3805 -----GATGACGCTCCAGCAGCTGCAGCTCAGCTCACTCCGACCTGG 3842
Db 2873 yTrpThrProTrpThrSerTrpSerSerCysSerGlnSerCysLeuAlaProGlyGlyGly 2893
Qy 3843 GGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3902
Db 2893 yProGlyTrpArgSerArgSer-Arg- 2901
Qy 3903 CACTTTGTGAGAGCAGCCCTGTATGCCAACACCTGTTGAGCCGAGACAGAGCTGTG 3962
Db 2902 --LeuCysProSer-----ProGlyAspSerSerCys- 2911
Qy 3963 AGCCTCTGGCCCTTCTGAGCGGCTGGCGTGTGATCGCATCGACCCCTGCTCTCTCC 4022
Db 2912 -----ProGlyAspAla-ThrGlnGluGluProCysSerProPro 2924
Qy 4023 CACCTCCCAAGGTCTACCGAGCTGGGAGGAGGTACAGTAGGCCCTGTCTCTGT 4082

Db 2925 -----IleGluCys----- 2927
Qy 4083 TCTACAGGAAGTCATGCTCGAGGAGTGTGAAGTGTTCAGGTTGTCAGAGGCGCTCA 4142
Db 2928 ---ThrGlyPheCysAlaProGlyCys-----Thr 2936
Qy 4143 TGGCTCTCT-----GCTTCTTGCCTACCACTTGGCAGTGGCCAGTGGCCACCC 4184
Db 2937 CysProProGlyLeuPheLeuHisAsnAlaSerCysLeuProArgSerGlnCysProCys 2956
Qy 4185 AGCCCTCAGTGGC-----ACATCTGGAGGCGAGGTTGAGGGG-----CACAC 4232
Db 2957 GlnLeuHisGlyGlnLeuTrpAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsn 2976
Qy 4233 CACACATCCCTTCTCTGGG-----GTGAAGCCCTTGGCTGCCCTCTCTCTTGA 4283
Db 2977 CysThrCysValSerGlyLysMetAlaCysThrSerGluArgCysProValAlaCysGly 2996
Qy 4284 TGGGTGTTGCTCCCTTATCCCCAAATCACTCTATACATCCAAATTCAGGAACAAACATGG 4343
Db 2997 Trp-----SerPro 2999
Qy 4344 TGGCAATTTCTACACAAAAGAGATGAGATTAACAGTGCAGGGTTCGGGTCTCTCATTTGGAG 4403
Db 3000 TrpThrLeuTrpSerLeuCysSerCysSerCysAsnValGlyIleArgArgArgPheArg 3019
Qy 4404 GTGCCCTATAAACACAGAGAGAAATACTGAAAGCACAGGCGGAGGAGCAGACC----- 4457
Db 3020 AlaGlyThrAlaProProAlaAlaPheGlyGlyAlaGluCysGlnGlyProThrMetGlu 3039
Qy 4458 -----AGACCAGCCAGGA-----GTCTCCAAAGCACAGAGTGGC 4493
Db 3040 AlaGluPheCysSerLeuArgProCysProGlyProValProGlyMetCysProArgAsp 3059
Qy 4494 AAACAAAACCCGAGCTGAGCATCAGACCT-----TGCTCGAATGTCT----- 4538
Db 3060 LysGlnTrpLeuAspCysAlaGlnGlyProAlaSerCysAlaGluLeuSerAlaProArg 3079
Qy 4538 ----- 4538
Db 3080 GlyThrAsnGlnThrCysHisProGlyCysHisCysProSerGlyMetLeuLeuVal 3099
Qy 4539 -----TCCAGTATTACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4568
Db 3100 SerProArgGlyHisProGlyProLeuGlyAlaSerValGlnProProValAlaLeuPro 3119
Qy 4569 CTTTCCCGAGGTATCTGGGTTCAGCTGGGAGG----- 4607
Db 3120 GlyAlaIleGlyThrGlySerValProGlyAlaGlyGlyTrpGlyProTrpGlyProTrp 3139
Qy 4608 -----CAA 4610
Db 3140 SerHisCysSerArgSerCysGlyGlyLeuArgSerArgThrArgAlaCysAspGln 3159
Qy 4611 CCATAGCCACACACAGGATTTCTGAAAGTTTCAATGACGATAGCATTTTGGGTGTAG 4670
Db 3160 ProProProGln----- 3163
Qy 4671 GTGGC-----AGCTCCCAAG----- 4688
Db 3164 --GlyLeuGlyAspTrpCysGluGlyProArgAlaGlnGlyValCysGlnAlaLeu 3182
Qy 4689 CCCTCCCTCC-----CCAGCCCAACCC 4709
Db 3183 ProCysProValThrAsnCysThrAlaIleGluGlyAlaGluTy-SerProCysGlyPro 3202
Qy 4710 ACTCATGACTTAAGTGTGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 4769
Db 3203 ProCysProArgSerCysAspLeuValHisCys-ValTrpArgCysGln-----Pr 3220
Qy 4770 TGATATTTATTCAGAAATTTCTTCTTCTTATTAACAATAAATGCTTGGCCCCAGAACT 4829
Db 3220 oGlyCysTrpCysProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnPr 3240

```

QY 4830 TAGT-----CTCTTGGCCAGCCTCAC-----4851
Db 3240 oGlyHisCysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGlyAlaArgLe 3260
QY 4852 -----CCCTCTGTGGTCACTAGACTCTTGGCCACCCCTGG-----4887
Db 3260 uAlaArgProAspGlyCysAenHisCysThrCysLeuGluGlyArgLeuAenCysThrAs 3280
QY 4888 -CTCCACCTCCCTGGCTGC-----CTCTG 4910
Db 3280 pLeuProCysProAspCysGlyGlyGlnSerLeuHieProCysGlyGlnProCysPr 3300
QY 4911 GTGAGCTGCACAGAGCTCTGGGAAGAGGCCCTCTCTCCCTCCCGCA-----4956
Db 3300 oArgSerCysGlnAspLeuSerProGlySerValCysGlnProGlySerValGlyCysGl 3320
QY 4957 -----CTGGGGCGATGGCGCACCTCAGACTT-----AC 4985
Db 3320 nProThrCysGlyCysProLeuGlyGlnLeuSerGlnAspGlyLeuCysValProAl 3340
QY 4986 CCACTGCTGCTGCACACCAACCC-----CTGATCCCTCAGTCTCCACACAGCTT 5039
Db 3340 aHisCysArgCysGlnTyrGlnProGlyAlaMetAlaProSerPheValPro-----SerTh 3359
QY 5040 CTGT 5043
Db 3359 rCys 3360

RESULT 5
ID Q9SPM0 PRELIMINARY; PRT; 1315 AA.
AC Q9SPM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Extensin-like protein.
GN PEX2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RX MEDLINE=21330248; PubMed=11437249;
RA Stratford S., Barnes W., Hohorst D.L., Sagert J.G., Cotter R.,
RA Golubiewski A., Showalter A.M., McCormick S., Bedinger P.;
RT "A leucine-rich repeat region is conserved in pollen extensin-like
RT (Pex) proteins in monocots and dicots.";
RL Plant Mol. Biol. 46:43-56(2001).
DR EMBL; AF159297; AAD55980.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR002965; P_rich_extensin.
DR Pfam; PF00560; LRR; 3.
DR PRINTS; PR01217; PRICEXTENSIN.
DR SMART; SM00370; LRR; 3.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
SQ SEQUENCE 1315 AA; 134402 MW; 64C97A2A01F0936F CRC64;

```

```

Alignment Scores:
Pred. No.: 9, 16e-20 Length: 1315
Score: 397.50 Matches: 301
Percent Similarity: 32.88% Conservative: 88
Best local Similarity: 25.44% Mismatches: 388
Query Match: 4.21% Indels: 407
DB: 10 Gaps: 52

```

US-09-931-704-3 (1-5087) x Q9SPM0 (1-1315)

```

QY 40 TCGCGGAGCGCGCTCGCCCTCCACTCCCGCAGCCTCCGGGAGGAGGAGCGCCACCCG 99
Db 461 AlaProAlaProMetArgMetProThrLeuArgSerProAlaAspGluTyrIlePro 480
QY 100 GCGGCGCCAGCCCGAGCCCGCATGGACCTCCG-----AGCAGGTTGAAACCCCAACTAG 153
Db 481 ThrProValProAla-LysSerProGlyThrSerProAlaSerArgGlyAl 500
QY 154 CCCTGCTCTTATACATGACAGCAGCGCCCATCTGATACCT---AAACCGACCAAGT 210
Db 500 aProLeuGlnAlaGlnProProAlaAlaSerPro-ProAlaThrProValLys 520
QY 211 CACAGCCCTCCA-----ACTCACCTCTGCTGCTGCCAGACC-----TCACCA 252
Db 520 er-SerProProAlaAlaValValLeuProProAlaLysThrProSerProPro 539
QY 253 CATCTTTGTGACTCAAACT-----CAACCGCTAAATCAACCAATCCCAAGTCTA 306
Db 540 AlaProValAlaSerProProGluAlaPro-ValSerSerProGlnProGlnVal-- 558
QY 307 AACTAATCTGAACCTTTTAAAGTAACCCAGCTCTTAAACCTAACCTAGCCCAATGCAAT 366
Db 559 -----LysSerPro-----ProProAlaProVal 567
QY 367 TATATCTACCTAGCCAAACCTTAAGTCTTGGCAGTCCAAAGTGTCCACTGAATCCT 426
Db 567 lAlaSerProProProMetLysSerProProProProAlaArgVal-AlaSerProP 587
QY 427 CACCTTGTCTCTACTGAAATCCAGAAAGCATATTTCCCACTGCCACATCCCTCC 486
Db 587 roProLeuMetLysSerProProProAlaProValAlaSerProProGlnProLeuL 607
QY 487 TTACAGCACCCCAACCCCTGCTCG-----ACTCTGGTATCTCTGGGATGTGCCAACT 540
Db 607 ysSerProProProValLeuTrpLeuSerThrProSerValLysSer---ProProP 626
QY 541 CTGAGTGCCATCAGCAACCAAGCCGACTCGTCAAT-----GCACCTC 585
Db 626 roProValProValAlaSerProProProProValLysSerProProLeuAlaProV 646
QY 586 TCTGCTTCTGTCTC-----CCACCTTGCAGGCTGATGAAGGCTCA 630
Db 646 alSerSerPro-SerProProValLysLeuProProLeuProAlaProGlyLysSerThr 665
QY 631 TTGAAGTCCAACCTTTTCCCACTTAACCAAGAGCGGGTGAACCTCCACACTGCCACC 690
Db 666 ProProProGluGluGluLysPro-ThrProProThrProValLysSerSerProProP 685
QY 691 GTTCCCTGAGAGTGAGCACTAAATCTCTTCAATCTAACCCCACTACACTTCCACAC 750
Db 685 o-----GluLysSerLeuProProProThrLeuThrSe 697
QY 751 TCAGGAATCACATCTAGAAATATACCCAAATAAGCCCAATAAGCAGCCCGACCTAG 810
Db 697 rProProProGlnGluLysProThrProSerThrProSerLysProProProProSe 717
QY 811 TGGTCTAACCTTATACCTTGTCTTATGGGTGAGTGTCTTCTGGCGCGCTCTCTC 870
Db 717 rPro-----ValGluThrLeuProProProSerLy 727
QY 871 CTGCTTCTCTCTTAGACTGACTGTGTCTAGCTGCCAGCTGTGACATGTGCTGTCTCC 930
Db 727 sSerSerProProGluProValSerSerProProGlnAlaProLysSerSerSerPr 747
QY 931 CACCTCTGACTCCCTCAAGCTGCAGTGGGACTGGAGACTGGAGAGCTGGCAGGAGCTAG 990
Db 747 o-----ProAlaProValSerSerProProProLeuLysSerSer-----760
QY 991 AACTGGAACACAGCGAGTGCACCTGAGTCCCTAGGCTGGCCCGCTCCCTCCATGTAC 1050
Db 761 -----ProProProValProGluSerSerProProPro-----771
QY 1051 ACATATATATGTTGGCACACACAGTGGCACACATGCCCAAGACTCTCTAGCTGAC 1110

```

```
Db 772 -----Thr- ProLysSerSerProLeu 779
Qy 1111 ACAGATCCATTCTCAAGTATCTACTAGACACTCATCGGTGCCAAGTCTCATCT 1170
Db 780 AlaProValSerProProGlnValGluLysThrSerProProAlaPro----- 797
Qy 1171 CAACATACACATGCTCTCTCTCTCTCCCTCTTCCAGGAGTGTTCCTCCCTCTCCAT 1230
Db 798 -----ValSerSerProProThrProLysSerSerProProLeuAla 812
Qy 1231 CCCCTGCTCCCTCCATCTGGTGTGTC-----CACCC-----TCA 1263
Db 813 ProValSerProProGlnValGluLysThrSerProProProAlaProValSerSer 832
Qy 1264 CCCCCCAGCCCAAGTGGGACACACACTGAGGGGTGCCAGTGTCTCCCGTGG 1323
Db 833 -ProProProThrProLys-----SerSerProProLeuAlaProValSerPro-- 849
Qy 1324 TGGGCCCCGGCGGCTCATGCTTCTCTCCATCTCTGCCACAGGGGACTCGTGGGGAT 1383
Db 849 ----- 849
Qy 1384 GTTAGCGTGTGCGACGGTGTCTGTGGACCTCTCCCTGAGTGCCAGCTCTCAATCGCAC 1443
Db 850 -----ProGlnValGluLysTh 855
Qy 1444 AGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCACCGGTACTCGGA 1503
Db 855 rSerProPro--Pro-AlaProValSerSerProProLeu--GluProLysProSerS 873
Qy 1504 GCACCACTCCGCACTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGGGAGT 1563
Db 873 erProProSerSerValSerSerProProThrVal----- 885
Qy 1564 TGGGAGGAGTGAGGAGTTGGGGAAGACAGTCTTAACCGTGGAGGGTCTGGTAATGA 1623
Db 886 -----LysSerSerPro----- 889
Qy 1624 TGGGAGGAGGGGCTCTTTGGCTCCACAGTCCCTCCCTGAGTGCCAGCTCTCATCTCGTCC 1683
Db 890 -----ProProAlaProLeuSerSer--ProProMet 900
Qy 1684 TTCCCTCTTAGTGGCCGCCCTTCCCATCCCTGCGCCAGGACTAGGCATGTGGGC 1743
Db 900 hrProLysSerSerProProAlaHisValSerSerProProGlu-----AlaG 917
Qy 1744 AGGCTCGCACCCGCTTGGCCCATTTGCCCATCTGGGTGCCAGCCAGCCGCCGCTCC 1803
Db 917 luLysSerSerPro-----ProLeuAlaProIle-----Ser-SerProProSerGlu 932
Qy 1804 CCTGGGGCCGGGAGTCTCTCTCTGTTTACACCGTGTGTGGTCTCTTTCGGGGG 1863
Db 933 ProLysSerProSerSerProProMet-ValGluLysThrSerProProAlaThrVa 952
Qy 1864 GGGGTGGGTGGGACAGAGGGGCCCACTCCCATGCTCGGCTTCCAGCTCGCTCTGC 1923
Db 952 1-----SerSerProProProThrPro-LysSerSerProProProA 966
Qy 1924 CCCCAGACTGGGGCCCTGCTGCTGTGGACCCAGGGGCTCCCTTCCTCGCTCTCC 1983
Db 966 laProValSerSerPro-----ProProValValLysSerS 978
Qy 1984 ATCTAGCTGGGCTCTAGGGGGTCAATGGGGAGGGGACTGTAGGGACCCAGGCAG 2043
Db 978 erPro----- 979
Qy 2044 TAGTGGCAGGGGTTTAGGGTGTGGATGGAGGTTATGCTGTAAAGATTGGGGGTGGTCC 2103
Db 979 ----- 979
Qy 2104 AGAGGTGTTTACAGAGCCCGAGGAGAGAGAGAGGGGTGGAGGAGCGACCAATG 2163
|||
Db 980 -----ProProAlaProValS 985
Qy 2164 GGGAACGGGCCCTCTTCCGTGTCTCTTCCACATCCAGACCTCTACTCTGGAGCCA 2223
Db 985 erSerProProProThrProLys--ProLeuProProAlaProValSerSer---- 1002
Qy 2224 GGGAAAGAAAGAGAGGTGGGGGAGCTGGTCCAGCCAGGATACACCGAGG 2283
Db 1003 -----ProProProValValLysSerSerProProProPro----- 1015
Qy 2284 AAATTAGTTGTCTCTGTGCTTGTGAGCGGTGAAACCTCCCTGGGCCCTTGCCTATCC 2343
Db 1016 -----ValSerSerProProProProProProProProProProProP 1027
Qy 2344 CAGGCTCTCCCTCTTCTTCTTCTTCTTCCAGTTATACATCTCCCTCATCTCTTCC 2403
Db 1027 ro--ProProProThrProValSerSerProProThr**LysProLeuProProPr 1046
Qy 2404 CTGGGCCCCAGCGCTCCCGAGGGTTGAAAGGGCTCTGCCCTCTTCCCTATACCATG 2463
Db 1046 oAlaProValSerSerProProProValValLysSerSerProPro--ProAlaProVa 1065
Qy 2464 CTGCTTCCATAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2518
Db 1065 lserLeuProProProThrArgLysProSerProProArgThrArgValSerSerProAr 1085
Qy 2519 -----TTCTGCAACCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2568
Db 1085 gProValValLysCysCysProProPro-----ThrLeuValSer-SerProProP 1102
Qy 2569 AGTACCG--TCTCTCCCGAGCCCTCAGCTGGTGGCGCTGGGTGTCACCGGCAAT 2625
Db 1102 roAlaProLysSerLeuProProPro----- 1111
Qy 2626 GGGCTCTGGTTCATCGGGCACTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2685
Db 1112 -----ProValSerSerProProProProProProProProProProProPro 1124
Qy 2686 TGCTTCACTCCACTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2745
Db 1125 -----ProProThrProValSerSerProProPro----- 1134
Qy 2746 AAATTTCTCAAGAGTGGTCTACACCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2805
Db 1135 -----AlaProLysSer--SerProProProThrProValSerSerProProGlu 1152
Qy 2806 CTTAAC-----CCCTGCAATCTGGCTTCCAGGCCCGCAGCAATGTTCTCTCCAAG 2856
Db 1152 uLysSerSerProProProAlaProValSerSerProProSer-----AlaProLy 1169
Qy 2857 GTCTGTCAGGCACCTCTCTTCCCAAGCCGACAGTGTGTTGAAGGCTCATCTCTCTCT 2916
Db 1169 sSerSer--ProProProAlaPro-----ValS 1178
Qy 2917 CTGTTTTCAGCCACACTGCTGAGCGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2976
Db 1178 erLeuProProProGluVal-----LysSerSerProProProAlaProI 1193
Qy 2977 TGCATCTCTCTGGGCCACTTCTACTCTCTCAGCTCTCTCAGGCTCTCTCTCTCTCTCT 3036
Db 1193 leSerSerProProProAlaLysSerProProProProAlaProMetSerSerLeu- 1212
Qy 3037 CCTGCCCCCAGAGCGGCACTCTCCCAAGTGTGTCGCCAGCCCAATCAGCAGCTCTCT 3096
Db 1213 -----ProProProValLysSerProProP 1221
Qy 3097 CTGAGCGTGTGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3146
Db 1221 roProAlaProValSerProProProProProMetLysSerProProProAlaProI 1241
Qy 3147 TGGAGAGCTCACCGGCACTGCTCTCACTGTCACCTGCATCAAAATGATATCTCTATT 3206
|||
Db 1241 le---SerSerProProPro----- 1246
```

```
QY 3207 GGAAGAACTCAGGAGGCCATGAACAAAGAGCCCTAGCATGGAGACAGGGCCAGTGTCTAG 3266
Db 1247 -----AlaProVal----- 1249
QY 3267 GGGACACAAAATAGAACTTTGGGAGCAGGTATCTCTTGGTGGAGCCAGCGGCTC 3326
Db 1250 -----LysProProSerL 1254
QY 3327 TGCCCTCTCTCCCATCCCTCCCTTTTACAGCTGAACCTACCTGGGCGCCCTT 3386
Db 1254 eu-ProProProAlaProValSerProProAla---ValThrSerAlaProPro 1272
QY 3387 TCAACGAGCCAGACTTCAACCTCCCTCC 3413
Db 1273 LysLysGluGluAapSerThrAlaPro 1281

RESULT 6
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Proline-rich mucin homolog.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=20090472; PubMed=10627046;
RA Espitia C., Lacleite J.P., Mondragon-Palomino M., Amador A.,
RA Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RT new family of fibronectin-binding proteins?";
RL Microbiology 145:3487-3495 (1999).
DR EMBL; AF071081; AAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01582; KV33CHANNEL.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8 CRC64;

Alignment Scores:
Pred. No.: 9,51e-20 Length: 763
Score: 396.50 Matches: 291
Percent Similarity: 32.86% Conservative: 77
Best Local Similarity: 25.98% Mismatches: 313
Query Match: 4.20% Indels: 441
DB: 2 Gaps: 59

US-09-931-704-3 (1-5087) x Q9XDH2 (1-763)
QY 1209 AGGAGTGTTCCTCCCT-----CCTCCATCCCTCTGCTCCCTCCATCTGGTGTCCC 1256
Db 8 ArgAlaLeuAlaProLeuProProAlaProProAlaProAlaProAlaGlu-ProLysSerLysPr 27
QY 1257 ACCTCACCCCCACCCAGCCCAAGGTGGGGACAGACACCTGAGGGGCTGCCAGCTGCTT 1316
Db 27 oProPheProAlaProProAlaPro----- 36
QY 1317 CCCCGTGTGGCGCGCGCGCTCATGCTTCTGTCATCTCCCTGCCACAGGGGACTCGT 1376
Db . 37 -ProCysTrp-----MetLeuVal----- 42
QY 1377 GGGGATGTTAGGCTGTGTGACGCTGCTGGGACCTCCCTGCGAGTGCAGCTCTCA 1436
Db 43 -----SerAlaAlaProProCysProProAlaProPr 53
```

```
QY 1437 ATGCACAGGGGACCCAGGGCTGG-----CCCCTCATCCAGA 1475
Db 53 oAlaProProLysProLysSerLysAlaProPheProProValProProAlaProAla 73
QY 1476 AAACCTATGACCTACCCGCTACCTGGAGCACCACCACTCCGAGCTTGGCTGGGACCTATG 1535
Db 73 aArgGluLeuAlaProProLeuProProAlaProPro----- 85
QY 1536 TGAGTATCCAGGTAGGATCTGGGAGTGGGGAGGAGTGAGGAGTGGGAAAGACAGT 1595
Db 86 -GluAlaPro-ArgGluSer----- 91
QY 1596 CTAACCCGTGGAGGCTTCTGGTAAATGATGGGTGAGGAGGGCTCTTTGGCTCCACCA 1655
Db 92 -----ArgProAlaLeuProProCysP 99
QY 1656 GTCCCTCTGTGTCTATCTCTCCCTCCCTCTTAGTGGCCCCCCCCACTTCCCAT 1715
Db 99 roProProValValIleProAspPro-ProGluProAlaAlaProProValPro--- 117
QY 1716 CCCTGGCCCCAGGACTAGGCATGTGGCAGGCTCGCACCCGCTTGGGCCATGGCCCA 1775
Db 118 ProAlaPro-----AsnSerProProPheProProPhePro 129
QY 1776 CTGGCTGCCAGCCAGCCGCGCTCCCTGGGGCGG---GGGAAGTCTCTCTGTT 1832
Db 130 ProAla-ProLysPheValProAlaProProValProProValProAsnSerProProPh 149
QY 1833 TACACCGTGTGTGTCTCTTTCGCGGGCGGGTGGGTGGGGACAGAGGGGCCCCAC 1892
Db 149 eProPro----- 151
QY 1893 CTCCTCCCTCGCTTCCAGCTCGCTCTGCCCCAGACCTGGGGCCCTCTCTCTGGA 1952
Db 152 -----PheProAlaAlaLeuAs 158
QY 1953 CCCAGGGGCTCCCTTCCGCTCTGCTCTCCATCTAGTGGGCGCTCTAGGGGGTCTCAT 2012
Db 158 nProProAla-----ProProAlaProProLeuAlaAsnSerProProLeuProAl 176
QY 2013 GGGGAAGGGGACTGTAGGAAACCCAGGAGTAGTGGCAGGGGTTTGGGTGTGATGG 2072
Db 176 a-ProProThrProAlaGlyThrProProAlaAla-----ProTrp----- 189
QY 2073 AGTTATGCTCTAAGGATTTGGGGTGTGTCCAGAGTGTTCAGAGGCCAGGAGAGAG 2132
Db 190 -----ProProVal-ProAlaAlaProLysSerLys 199
QY 2133 GAAGGAGGTTTGGAGGCCAGGACCCATGGGGAAACCGGCCCTCTTCCCTGTGT--- 2189
Db 200 ProAla-----SerProProArgProProAlaProProMetProAlaThrProMetGlu 217
QY 2190 -----CCTCTTCCATCCAGACCCCTACTCTGTGGAGCCAGGAAAGAAAGGAGAAG 2243
Db 218 PheProProLeuProProValProPro-----AspPro 228
QY 2244 GTGGCGGGGAGCTGGCTCCAGCCCGGAGATACACCGAGGAATTAGTTGTCTCTGTGC 2303
Db 229 IleSerLysGluThrProProAlaPro----- 237
QY 2304 TTGTACAGCGTGTGAACCTCCCGCTTGGGCGCTTTCCTATCCAGGCGCTCTCCCTTCTTC 2363
Db 238 ---ProAlaProProIleProProAlaProValProIlePro---ProValProProLeuP 256
QY 2364 TCCCTTCTTTCCAGTTATATCTCCCTCATCTTCCTTCCCTGGGGCCCGAGCGCTCCCC 2423
Db 256 roPro-----ValProAsnLysIleProProAlaProAlaProP 270
QY 2424 CGAGGGTTGGAAGGGCTCTGCCCTTCTCCCTATACCATGTGTCTTCCATAGCTTCTCT 2483
Db 270 roValAla-----ValAlaAlaValLeuValAlaProCys-----Pro-ProLeuPro 285
```


Db 358 -----IleGlnGlyHisProProGlySerLeuTyProProAsnSerValLeu 374
Qy 1025 AGGCTGGCCCGCTCCATGATACACATATACATGTTGGACACACACAGTGGCAC 1084
Db 375 AlaProGlySerIleProPro--LeuArgProSerAsnGlnGlyHisProProGlys 394
Qy 1085 AC---ATGCCAAGACTCTCTCAGCTGACACACACATCCATCTCAAGTATCTACTGATA 1141
Db 394 erLeuLeuProProAsnArgAlaLeuPheArgLeuLeuAspHisGlnIleLeuAspA 414
Qy 1142 GACACTATGCTGCCAAGTCTCATCTCAACATACACATGCTCTCTTTCTCTCCG 1201
Db 414 snHis-----GlnAspHisCysLeuLeuMetTrpGluT 426
Qy 1202 TCTTGCAGGAGTGTTCCTCCCTCTCCATCCCTCTGCTCCCATCTGGTGTCCACCT 1261
Db 426 yrIlePro-----HisPro-----ArgAlaGlyPheProProS 437
Qy 1262 CACCCCCACCCAGCCCAAGTGGGACACACACCTGAGGGCTGCCAGCTGCTTCCCG 1321
Db 437 erGlyProLeu--ThrGlnGlyGly-----HisProThrGlySerLeuLeuProProAs 454
Qy 1322 TGTGGGC-----CGGGCGGCTCATGCTTCTCTGTCATCTCTGCCACAGGGGACTCG 1375
Db 454 nThrGlyLeuProProGlySerIleProProLeuArgProPro----- 468
Qy 1376 TGGGGATGTAGCTGCTGTGCAGGTGTCTGCGACCTCCCTGCGAGTCCACTCTC 1435
Db 468 ----- 468
Qy 1436 AATCGACAGGGGACCCAGGGCTGGCCCTCCCA---TCCAGAAAACCTATGACCTCAC 1492
Db 469 -----AsnGlnGlyGlyHisProProGlySerGlnLysPro----- 480
Qy 1493 CGCTACTGGAGCAACAACCTCCGAGCTGGCTGGGACCTATGTGAGTATCCAGCGTAGG 1552
Db 481 -----ProAsnValG 484
Qy 1553 AATCTGGGAGTTGGGAGGAGTGAGGAGTTGGGAAAGACAGCTCTCAACGTTGGAGGTT 1612
Db 484 ylleTy-----ProProSerThrGly----- 491
Qy 1613 CTGGTAAATGATGGGTGAGGAGGGCTCTTTGGCTCCACCCAGTCTC----- 1658
Db 492 -----TrpIleProProPheGlyProLeuThrG 501
Qy 1659 -----CCCTGCTGTGCTATCTCT-----GCCCTTCCCTTTAGTGGC 1699
Db 501 nGlyGlyHisProProGlySerLeuLeuProProAsnThrGlyLeuProProGlySerI 521
Qy 1700 CCCCCCATCTCCATCCCTGGCCCGGAGGACTAGGATGGGAGGCTCGCACCCGCG 1759
Db 521 eProProLeuGlySerPro-----AsnGlnIleGlyHisProPr 534
Qy 1760 TTGGC---CCATTGCCCACTGGCTGCCAGCCAGCCCGCCCTCCCTCGGGGCGCG 1816
Db 534 o-GlySerGlnLysProProAsnValGlyIleTyProProSerThrGlyTrpThrProp 554
Qy 1817 GGAAGTCTCTCTGTTTACACCGTGTGGTGTCTCTTGGCGGGGGGGTGGTGGG 1876
Db 554 roSerGlyProLeu----- 558
Qy 1877 GACAGAGGGGCCCCACCTCCCATGCTCGTGTCCAGCTCGCTCTGCCCCAGACCTGGG 1936
Db 559 -----ThrGlnGlyHisProProGlySerLeuLeuProProAsnThrG 574
Qy 1937 GCCCTGCTCTGTGGACCCAGGGGCTCCCTTCCGCTCGCTCTCCATCTCTAGCTGGGC 1996
Db 574 ly-----LeuProProGlySerIleProProLeuGlySerProAsnGlnIleGlyH 591
Qy 1997 CTCCTAGGGGGGTATGGGGAGGGGAGTGTAGGG---AACCCAGGAGTGTAGTGCAGG 2053
Db 591 isPro-ProGlySerGlnLysProProAsnValGlyIleTyProProSerThrGly--- 609

Qy 2054 GGGTTTAGGGTGTGATGAGGTTATGCTGTAAGGATTGGGGTGGTCCAGAGGTGTC 2113
Db 610 -----TrpIlePro-----ProSerGlyPro-----Leu 617
Qy 2114 AGAGAGCCAGAGAGAGAGAGAGGAGGTTGAGAGCGGAGGACCAATGGGGAAACCGGC 2173
Db 618 ThrGln-SerGlyHisProProGlySerLeuLeuProPro--AsnThrGlyLeuPro-- 635
Qy 2174 CCCCTCTTCCCGTGTCTCTCTTCCATCCAGACCCCTACTCTGGAGCCAGGAAAGAA 2233
Db 636 -ProGlySerIleProProLeuArgProProAsnGlnGlySer----- 649
Qy 2234 AGGGAAGAAGTGGCGGGGAGCTGGCTCCAGCCCGACATACACCGAGGAAATAGTTT 2293
Db 650 -----LeuLeuProProAsnThrGlyLeu----- 657
Qy 2294 GTCTCTGTGTGTGACGCTGTGAACCTCCCTGGGCGCTTGCCTATATCCAGGCTCTC 2353
Db 658 -----ProSerGlySerIlePro----- 663
Qy 2354 CCCTGTCTTCTCCCTTCTTCCAGTTATATATCTCCCTCATCTCCCTTCCCTGGGCCCA 2413
Db 664 -----ProLeuGlySerProAs 669
Qy 2414 GCGGCTC-----CCCGAGGTTGAAAGGCTCTGCCCTCTCCCTATATACATGCTGT 2467
Db 669 nGlnIleGlyHisProProGlySerGlnLysProProAsnSerGlyIleTyProProSe 689
Qy 2468 CTTCCATAGCTTCTCT-----CCTGCTCTATCATAGACTGCTCCA-----TT 2512
Db 689 rThrGlySerIleProProSerGlyProLeuThrGlnGlyHisProProGlySerLe 709
Qy 2513 TCTTCTTCTGCAACCTGCTCTCATCTCTCTCTGTTCTTTCGAGGTGTAGTGAGTA 2572
Db 709 uLeuProSerAsnThrGlyLeu----- 716
Qy 2573 CCGCTCTCTCCCGACCTCAGCTGGTGGGCTGGTGTGTGTCAGCGGCAAAATGGGGCTC 2632
Db 717 -ProSerGlySerIleProProLeuGlySerProAsnGlnIleGlyHisProProGlySe 736
Qy 2633 TGGTTCCAAATGGGCCACTCTCATCTCTCTCTGTTCTTGTGTCAGAAAACCTTGTCTCA 2692
Db 736 r-----GlnLysPro----- 739
Qy 2693 CTCCTACTGCTCTCTAGTTCCCGACCTTTTCTCTCTCTGCTTTCCTGCAAAATTC 2752
Db 739 ----- 739
Qy 2753 TCCAAGGAGTGTCTACACCTCTGCTCTCCTCTCTCTCCACCCACTCACTTCTTTAAC 2812
Db 740 -ProAsnSerGlyIleTyPro-----Pr 747
Qy 2813 CCCTCAATCTGGCTTCAGGCCCAGGCAATGGTCTCTCAAGTGTCTGAGGCTGTCAGCACCTCC 2872
Db 747 oSerThrGlyTrpIlePro--ProSerGlyProLeuThrGlnGlyHisProPro-- 765
Qy 2873 TTGCCAAGCCCGACAGTGTGTTGAAGGCTCATCTCTCTGCTGTCTGTTTGGCAGCACCA 2932
Db 766 -----GlySer----- 767
Qy 2933 CTGCTGAGGCTGTGCTCTCTCGAACTCTCTCTCTCTGCTCTCTGCACTCTCT---GG 2989
Db 768 -----LeuLeuProProAsnThrGlyLeuProProGlySerIleSerProLeuG 784
Qy 2990 GCCACTTCTACTCTCTCCAGCTCTCCAGGCTCTCTTCC-----TCTCTGTC 3037
Db 784 ySerProAsnGlnIleGlyHisProProGlySerGlnLysProProAsnValGlyIleTy 804
Qy 3038 CTGCCCCCAGCGGGGACTCTCCCAAGGTTTGGCCACCCAGCAGCAATCAGCAGCTCTTC 3097
Db 804 rProProSerThrGlySerIleProProLeuArgProProAsnGlnGlyHisPro-- 823

Db 602 ProLysCysTyrProGlySerSerAspProGluCysLeuAsnCysTyrProGlySerPro 621
:::
Qy 142 AACCCAACTAGCC-----TGCTCTTCATACATGAC 174
:::|||||
Db 622 AspProArgCysProLysValProThrThrLysLysAlaGlyCysTyrGluGlySerAsp 641
:::|||||
Qy 175 AAGCAGCC-----CCCATCTGATACCTAAACACCAAGTACACAGCCCTCCAACT--- 225
:::|||||
Db 642 AspProArgCysGlnProAlaThrThrLeuProProSerThrArgArgProProThrIle 661
:::|||||
Qy 226 CACCTCTGCCT-----GCCAGACCTCACACATCTCTGTGGACTCAAACT 273
:::|||||
Db 662 AlaProArgProArgCysTyrLeuGlySerThrAspPro---SerCysProGlnGlnThr 680
:::|||||
Qy 274 CAACCGCACTAAATCAACCAATC-----CCAGTCTAAAC 309
:::|||||
Db 681 GlnProThrThrArgThrProIleThrThrIleLysProArgCysPheProGlySerAsn 700
:::|||||
Qy 310 TAATCTGAACCTTTAAAGTAACCCAGTCTCTAAACCTAAACCTAGCCCAATGCCAATTAT 369
:::|||||
Db 701 AspProGluCysGlnProAlaThrThrArgArgProLeuThrThrSerLysProAsn--- 719
:::|||||
Qy 370 ATCTACCTAGCAAACTCACTGCTGCTGTCAGTCCAAAGTGTCCACTGAATCTCTCAC 429
:::|||||
Db 720 CysTyrProGlySerThr-----AspArgArgCysProGlnGluProVal 734
:::|||||
Qy 430 CTGTGCTCTCACTGAATATCCAGAAAGCATATTTCCCACTGCCACATCCCTCTTA 489
:::|||||
Db 735 Thr-----ThrSerArgProArgCysTyrProGlySerSerAspProGluCys 750
:::|||||
Qy 490 CAGCACCAACCTCGCTCTGACTCTCTGCTGATCTGGGATGTCCAACTCTGCAG--- 546
:::|||||
Db 751 GlnProAlaThrLeuValProProThrThrValArgThrThrIleThrThrSerArgGln 770
:::|||||
Qy 547 -----TGCCATACAGCCAAAGCCGCTCGTCAA 576
:::|||||
Db 771 ArgCysTyrProGlySerSerAspProSerCysGlnProProThrThrArgLysProIle 790
:::|||||
Qy 577 ATGCACTCTCTCCCTTC-----TGTCCTCCACCTTCG 609
:::|||||
Db 790 eThrThrProLysProAsnCysTyrProGlySerThrAspSerArgCysProGlnThrPr 810
:::|||||
Qy 610 CAGGCTGATGGAAGCTCTCATTAAGTCCA-----ACTTTTCCCACTAACACCAAG 663
:::|||||
Db 810 oArgThrThrGlnGlnProArgCysTyrProGlySerThrAspProAspCysGlnProAl 830
:::|||||
Qy 664 AACGGGTGAACCTCCACACTGCCCGTTCCTCGAGTGAAGTAAATCTCTCTCAA 723
:::|||||
Db 830 aThr----- 831
:::|||||
Qy 724 TCTAACCCCACTACCTCCACACT-----CAGGAATCACATCTAGATATACCC 777
:::|||||
Db 832 -TyrLeuProProThrThrAlaArgThrThrIleThrThrAlaLysProArgCysTyrPr 851
:::|||||
Qy 778 AAAAATAAGC-----CCATAAGGAGCCGACCTAGTGTCTA-----ACCTATACCTTGC 831
:::|||||
Db 851 oGlySerSerAspProTyrCysGlnProProThrThrLysLysProAsnCysTyrProG 871
:::|||||
Qy 832 TTCCTATGGGTGAGTGTCTTGGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 891
:::|||||
Db 871 ySerThrAspSerArgCysProGlnLysProIleThr-----Pro-LysProL 888
:::|||||
Qy 892 ACTGTGCTCAGCTGCAGCTCTGACATGCTGTCTC-----CCACCTCTGACT 942
:::|||||
Db 888 yCys---TyrProGlySerThrAspLeuGluCysLeuAsnCysTyrProGlySerProA 907
:::|||||
Qy 943 CCCCTCAAGCTGAGTGGGAGCTGGAAGTGGCAGGAAGTGGGTACAACTGGGAACA 1002
:::|||||
Db 907 spProArgCysPro---LysVal-ProThrThrLysLysAlaGlyCysPheAspGlySer 925
:::|||||
Qy 1003 GGCAGTGCAGCTCAGTCCCTAGGCTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1062
:::|||||

Db 926 AspAspProLysCysGlnProAlaThrThrTyrLeuProProSerThrLysArg----- 942
:::|||||
Qy 1063 GTTGGCACACACAGTGGGCACACATGCCAAAGACTCTCTCAGCTGCACACAG----- 1116
:::|||||
Db 943 -----ThrProThrIleAlaProLysProArgCysTyrProGlySerLysAspProGly 960
:::|||||
Qy 1117 -----ATCCATTCTCAGTATCTACTATGATACACTCATGCGTGCCA--- 1158
:::|||||
Db 961 CysProGlnIleThrThrArgAlaProThrSerThrSerArgProArgCysTyrProGly 980
:::|||||
Qy 1159 AGTCTCATCTCTCAAAACATACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215
:::|||||
Db 981 SerThrAspProGluCysHisProThrThrSer---SerProAlaIleThrArgIlePro 999
:::|||||
Qy 1216 -----TTTCCCTC-----CTCCATCCCTCTGCTCTCCATCTGGTGCCCA 1257
:::|||||
Db 1000 ValThrThrArgIleProLeuThrThrAlaLysProArgCysTyrProGlySerGlnGlu 1019
:::|||||
Qy 1258 CCC-----TCACCCCCCAACCCAGGCTGGGAGCAGACACCTGAGGGCTGCCAGC 1311
:::|||||
Db 1020 ProGlyCysGlnProAlaThrArgProIleThrThr-----SerLysProArg 1036
:::|||||
Qy 1312 TGCTTCCCGTGTGGGCGGCGCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1371
:::|||||
Db 1037 CysTyrPro----- 1039
:::|||||
Qy 1372 CTGCTGGGGATGTAGCGTGCCTGTGCACCGTGTCTGGCACCTCTCCCTGAGTGCACG 1431
:::|||||
Db 1040 -----GlySerLeuAspProGluCysGlnProSer 1049
:::|||||
Qy 1432 TCTCAATGCACAGGGGACCCAGGCGCTGGCCCTCTCCATCCAGAAACCTATGACCTCAC 1491
:::|||||
Db 1050 ThrTyr----- 1051
:::|||||
Qy 1492 CCGCTACTGGAGCACCACTCCGAGCTTGGCTGGACCTATGTAGTATCCAGCGTAG 1551
:::|||||
Db 1052 ---LeuPro-----ProThrPro----- 1056
:::|||||
Qy 1552 GAATCTGGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGT 1611
:::|||||
Db 1057 -----ValArgThrThrValProThrIleProThrThr-ArgIle 1069
:::|||||
Qy 1612 TCTGTAAATGATGGGTGAGGAGGCTCTTTGGCTCCACAGTCCCTCTCTCTCTCTCTCT 1671
:::|||||
Db 1069 eProValThrThr-----SerLysProAsnCysTyrProGlySerThrAs 1084
:::|||||
Qy 1672 TATCT 1725
:::|||||
Db 1084 pArgArgCysProLys-GluProValThrThrProLysProArgCysTyrProGlySerP 1104
:::|||||
Qy 1726 AGGACTAGGCACTGGGAGGCTGCACCCGCTTGGCCCATTTGCCACTGGCTGCCA 1785
:::|||||
Db 1104 roAsnProGluCysGlnLysAlaThrTyrSerPro----- 1115
:::|||||
Qy 1786 GCCAGCCCGCT 1845
:::|||||
Db 1116 --ProThrThrArgThrProValThrThrSerLysProAsn----- 1128
:::|||||
Qy 1846 GGTGTCT 1905
:::|||||
Db 1129 --CysTyr-----ProGlySerThrAspSerArgCysProGlnLysPro----- 1142
:::|||||
Qy 1906 GTTCCAGCTCGCTCTGCCCCCAGACCTGGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1965
:::|||||
Db 1143 -----ProThrThrLeuLysProLysCysSer---ProGlySerSerA 1156
:::|||||
Qy 1966 CTTCCGTCT 2025
:::|||||
Db 1156 spProGluCysLeuAsn----- 1161
:::|||||
Qy 2026 TGTAGGAACCCAGGAGTGTAGTGGAGGGGTTTAGGGTGTGGATGGAGGTTATGCTGTA 2085
:::|||||
Db 1162 -----CysT 1163
:::|||||


```
Db 131 oHis----- 132
QY 339 CTTAAACCTAACCTAGCCCAATGCAATTATATCTACCTAGCCAAACCCCTAACCTGCTT 398
Db 132 ----- 132
QY 399 TGCAGTCCAAAGTGTCCACTGAATCTCCTGCTTGGTCTCCTCACTGAAATCCAGAAAG 458
Db 133 -----ProLysProThrLysProHis-----ProHisProLysProThrLys 148
QY 459 CATATTCCCACTGCCCAATCCTCTCTTACAGACCAAC-----CTGGCCTCTG 511
Db 148 sHisHisProHisProLysProThrLysProHisProLysProLysProSerVally 168
QY 512 GACTCTCTGGTATCTCTGGGATGTCAAACTCTGCAGTGCATCAGCAACAGCCGACTC 571
Db 168 sProPro-----ProSerThrProLysProProth 178
QY 572 GTCAATGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 631
Db 178 rThrAsnProProSerThrProGlnPro----- 188
QY 632 TGAAGTCCAACTTTTCCCACTTAACACCAAGAGGGGTGAACCTCCACACTGCCAC-- 689
Db 189 -----ProThr-----HisLysProProCysThrPr 198
QY 690 -----CGTTTCCTGAGAGTGAGCACTAAATCTCTCAATCTAACCACCC 736
Db 198 oThrProProValAlaSerProProMetAlaThrProProThrGlnMetProProIleAl 218
QY 737 TACACTTCCCACTCAGGAATCAGTCTAGAGATATACCAAACTAAGCCCAAGG 796
Db 218 aThrProProIleAlaLysSerProValAlaThrProProIleAlaThrProProThrAl 238
QY 797 CAGCCGACCTAGTGTCTAAACCTATACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 856
Db 238 aThrProProIleThrIleProProValAlaThrProProIleThrProProIleAl 258
QY 857 CGGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 916
Db 258 laAsnProProIleMetProProIleAlaThrProProValAlaAlaProProIleT 278
QY 917 CA---TGTGTCTCTCCACCTCTGACTCCCTCAAGCTGAGCTGGGAGACTG 973
Db 278 hrAsnProProIleSerLysPro-----ProVal----- 287
QY 974 GCAGGAAGCTAGGTACAACTGGAACAGCAGGCTGAGCTGAGTCTGAGTCTGAGCTGGC 1033
Db 288 -----ThrThrProProThrThrThrProProIleAlaLysP 300
QY 1034 CCGTCTCTCCATGTACACATATACATGTGGCACACACAGTGGCACACATGCAA 1093
Db 300 ro-----ProIleAlaThrProPro- 306
QY 1094 AGACTCTCTCAGCTGACACAGATCCATCTCAAGTATCTAGCTAGACACTCATGCG 1153
Db 307 -----IleSerThrProAlaAlaThrP 315
QY 1154 TGCCAAGTCTCATCTCAACATACACATGC-----CTCTCTTCT 1195
Db 315 roProAlaAlaThrProProIleThrThrLeuProProAlaLysProProValAlaIleS 335
QY 1196 CTCCTCTCTG-----CCAGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1246
Db 335 erProIleValThrProProValThrProIleAlaGlnProProValAlaThrProProT 355
QY 1247 CTGGTCTCCACCC-----TCACCCCAACCCAG----- 1275
Db 355 hrAlaThrProProValAlaThrProProIleAlaThrProProThrSerLysSerProI 375
QY 1276 -----CCCAAGTGGGGACAGACACTGAGGGGTGCCA----- 1309
Db 375 leSerThrProProIle-SerGluSerProValAlaThrProProThrAlaThrSerPro 394
```

```
QY 1310 -----GCTGCTTCCCGTGTGGGCGC 1336
Db 395 lleLysThrProProAlaLysProProValAlaThrProProIleAlaLysSerPro 414
QY 1337 CGCTCATCTTCTCTGTCATCTCTGCCACAGGGGACTCGTGGGGGATGTTAGCTGCTG 1396
Db 415 lleAlaThrProProThrAlaThrProPro-----Val 425
QY 1397 TGCAGGTGCTCTGCACCTCTCTGAGTGCAGCTCTCAATCGCACAGGGGACCCAGGG 1456
Db 426 AlaThrProProIleGluLysProProValAlaThrProProThrThrProPro- 445
QY 1457 CTGGCCCTCTCATCCAGAAACCTATGACCTCACCCGCTACCTGAGCACCACTCCGC 1516
Db 445 rAlaThrProProValAlaLysProProValGlnThrProProIleAlaThrProProTh 465
QY 1517 AGCTTGGCTGGACCTATGTAGTATCCAGCGTAGGAATCTGGGAGTGGGGAGGATGA 1576
Db 465 rAla----- 466
QY 1577 GGAGTTGGGAAGACAGCTCCTAACCGTGGAGGGTCTGTAATATGATGGGTGAGGAGG 1636
Db 466 ----- 466
QY 1637 GGCTCTTGGCTCCCAACAGTC-----CCCTGTCTGGTCTATCTCTGCCCTTCC 1687
Db 467 -----LysProProIleSerThrProProIleSerLysProProValAlaThrPr 483
QY 1688 CTCTTAGTGGCCCCCACTTCCCATCTCTGGGCCAGGACTAGGCATGTGGGAGGC 1747
Db 483 oProAlaAlaThrProProIleThrThrProPro-----VallysProProVa 500
QY 1748 CTGCACCGCTTGGCCCATCTGCCCTGCT----- 1781
Db 500 lAlaThrProProLeuAlaIleProProValAlaLysProProValValThrProProTh 520
QY 1782 -GCCACCGCAGCGCCCTCTCTCTGGGGCGC---GGGAAGTCTCTCTCTTTACAC 1837
Db 520 rAlaThrPro-ProIleAlaThrPro-----ProIleAlaLysSerProValAlaThrP 538
QY 1838 CGTGTGTGTGTCTCTTGGCGGGGGGGTGGGTGGGACAGAGGGGCCCACTTCC 1897
Db 538 ro-----ProThrAla 542
QY 1898 ATGCTGTGGT-----CCAGCTCGCTCTGCCCCAGACCT 1933
Db 542 hrProProValAlaThrProProIleAlaLysProProValValThrProProThrThrT 562
QY 1934 GGGGCTCTGCTCTGACCGCCAGGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1993
Db 562 hrProProThrAlaThrProProValAlaLysProProValAlaThrProProIleAlaT 582
QY 1994 GGCCTCTAGGGG-----GTCAATGGGGAAGGGGACTGTAGGG 2032
Db 582 hrProProThrAlaLysProProIleSerThrProProIle-SerLysSerProValAla 601
QY 2033 AACCCAGCAGTACTGGCAGGGGGTTTAGGTGTGTAGTGTATGCTGTAGGATTT 2092
Db 602 ThrProProAlaAla----- 606
QY 2093 GGGGTGTCTCAGAGGTGTTTTCAGAGAGCCAGGAGGAAGAGGGGTGGAGGAGCC 2152
Db 607 -----ThrProProIleThrThrProPro-----ProAla 616
QY 2153 GAGGCACATGGGAACCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2212
Db 617 LysProProValAlaThrProProIle-AlaThrProProIleAlaLysPro----- 633
QY 2213 CTCTGGAGCCAGGGAAGAAAGAGGAGGTGGGGGGGAGCTGGCTCCAGCCCGCAGG 2272
Db 634 -----ProValAlaThrProProTh 640
```


Db 257 ProProProGlyProAla-ProProGlyAlaArgProProProGlyProProProGly 276
Qy 2056 GTTATAGGTTGGATGAGGTTATGCTGAAGATTGGGGTGGTCCAGAGGTTTCAG 2115
Db 276 y----- 276
Qy 2116 AGAGCCAGGAGAGAGAGAGGTTGGAGGAGCGAGCCACCATGGGAAACCGCCC 2175
Db 277 ---ProProProProGlyProAlaProProGlyAlaArgProProGly-ProProP 295
Qy 2176 CTTCTCCGTTGCTCTCCATCCAGACCTACTCTGGAGCCAGGAGGAGAAAG 2235
Db 295 roProGlyProProPro---ProGlyProAlaPro---ProGlyAlaArgProProG 313
Qy 2236 GGAAGAGGTTGGGGGAGGCTGCTCCAGCCCGAGGATACACCGAGGAAATTAGTTGT 2295
Db 313 lyProProProProGlyProProProProGlyPro----- 324
Qy 2296 CTCTGTGTTGTCAGCGTGAACCTCCCTGGGCGCTTGCCTATCCAGGCTCTCCC 2355
Db 325 ---AlaProProGlyAlaArgProProProGlyPro---ProProProGlyProProP 342
Qy 2356 CTTGCTCTCCCTCTTCCAGTTATACATCTCCCTCATCCCTTCCCTGGGCGCCAGC 2415
Db 342 ro-----ProGlyProAlaP 347
Qy 2416 CGCTCCCGAGGTTGAAAGGCTGCTGCCCTCTTCCCTATACCATGCTCTTCCATA 2475
Db 347 ro-----ProGlyAla----- 350
Qy 2476 GCCTTCTCTCTCTACTATGAGACTGCTCCATTTCTTCTTCTGCAACCTGTCTCC 2535
Db 351 ---ArgProProProGlyProProPro----- 358
Qy 2536 TATCAGCTGAACCTTCTTTCGGAGTGTAGTAGTACCCGCTCTCCCGAGCCCTCAG 2595
Db 359 -----ProGlyProProProGlyP 366
Qy 2596 CTGGTGGCCCTGGGTGTCTCAGCGCAATGGGCTCTGGTTCCAAATGGGCACTCTCAT 2655
Db 366 roAlaProProGly----- 370
Qy 2656 CTCTCTTGTCTTGTGCAAAACCTTTGCTTCACTCCACTGCCCTCTCTAGTTCCC 2715
Db 371 -----AlaA 372
Qy 2716 GACCTTTTCTCTCTGCTTCCCTGCCAAATTTCTCAAGAGTGTCTACACCTC 2775
Db 372 rgPro-----ProProGlyProProProProProProProProGly-----ProA 388
Qy 2776 TGCTTCCACTTCTCTCCACCCACTCACTTCTTAACCCCTGCAATCTGGCTTCAGGCC 2835
Db 388 laProProGlyAlaArgProProProGlyProProProProGlyProProProGlyP 408
Qy 2836 CCAGCAATGGTTCTCCAAGTCTGTCAGGACCTCTTGTCCAAAGCCGACAGTGTGTTG 2895
Db 408 roAlaProProGlyAlaArgProProProProProProProAlaaspGluProGIng 428
Qy 2896 AAGGCTCATCTCTGCTGTCTGTTTTCAGCCACATGCTGAGCGGTGCTGCCCTCTCT- 2954
Db 428 lnglyPro-----AlaProSerg 434
Qy 2955 -----CGAACTCTCTTCTTCTGCTCTGCACTCTCTGGGCGCACCTT 2997
Db 434 lyAspLyProLysLyProProProProAlaGlyProProProGlyProProS 454
Qy 2998 CTACTCTCCAGCTCTCTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3057
Db 454 exProGlyProAlaProProProGlyAlaArg-----ProProProGlyProP 469
Qy 3058 CTCCCAAGTTTGGCCACCCAGCCCAATCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCT 3117
Db 469 roProProGlyProProGlyProAlaProProGlyAlaArgProProProGlyP 489

Qy 3118 CT 3167
Db 489 roProProProGlyProProProProGlyProAlaProProGlyAlaArgProProPro- 508
Qy 3168 TGCTTCACTGTCACTGTCATCAATATGATATCTTATTGGAAAACTCAGGAGGCGCAT 3227
Db 509 -----GlyProProp 512
Qy 3228 GAACAAAGAACCTTAGCATGGAGACAGGGCCAGTGTCTAGGGGACACAAAAATAAGAACT 3287
Db 512 roProGlyProProProProProProGlyProAlaProPro----- 523
Qy 3288 TTGGAGCAGGTATCTCTTGGTGTGAGCCAGCGGTCTGCCCTCTCTCTCTCTCTCTCTCTCA 3347
Db 524 --GlyAlaArgProProProProProProGly-ProProProProGlyProAla 542
Qy 3348 CCCTCA 3407
Db 543 Pro---Pro-GlyAlaArgProProProGlyProPro-----ProPr 555
Qy 3408 CTCCCGCTCTGGGGCAGAGACTCTGCCAGGCGCACTGTTGACTTGGAGGTGTGGCGAA 3467
Db 555 oGlyProProPro-----ProGlyProAlaP 564
Qy 3468 GCCTCAATGACAACTGGGCTGACCCAGAACTACGAGGCTACAGCCACTCTCTGTGT 3527
Db 564 oProGlyAlaArgProProProGlyProProProProGlyProProProGly----- 582
Qy 3528 ACTTGGCTGGCTCAACCGTCAGGCTGCCACTGTCTGAGCTGGCGCAGCTGGCCCACT 3587
Db 583 -----ProAlaProProGly-----AlaArgProProProGlyPro- 594
Qy 3588 TCTGCACCGCTCTCAGGGCTGCTGGGCGCACTTGGGCGCTCATGCGAGCTCTGGGT 3647
Db 595 -----ProProProGlyProPro----- 600
Qy 3648 ACCCACTGCCCGCCGCTGCTGGAGCTGAACCCACTTGGACTCTCTGGCTGCGCCACA 3707
Db 601 -ProProGlyProAlaPro-ProGlyAlaArgPro-----ProProGlyPro--ProPr 617
Qy 3708 GTGACTTCTCTCAGAGATGAGCACTTCTGCTGCTGAAGGAGCTGCGAGCTGGCTGT 3767
Db 617 oProGlyProProProProGly----- 624
Qy 3768 GCGCTCGGCCAAGCACTTCAACCGGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3827
Db 625 -----ProAla-----ProProGlyAlaArgProProProGly 635
Qy 3828 TCACCTGCACTGGGGCTCATGGCTTCTGACTTCTGACTTCTCTCTCTCTCTCTCTCTCT 3887
Db 635 yProProProProGly-----ProProPr 643
Qy 3888 TTCAACCT 3947
Db 643 oProGlyProAlaProProGlyAlaArgPro-ProProGlyProProProProProPro 663
Qy 3948 GAGACAGAGCTGTGAGCTCTGGCCCTTCTCTGAGCCGCTGCTGGCGGTGTGTGATGCA 4007
Db 663 ly----- 663
Qy 4008 GCCCTGTCTCTCCCACTCTCCCAAGGTCTACCG 4042
Db 664 --ProSerProProArgProProProGlyProPro 674

RESULT 11

Q9NYW9 PRELIMINARY; PRT; 1548 AA.
ID Q9NYW9
AC Q9NYW9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Somatostatin receptor-interacting protein splice variant b.

[illegible]

[illegible]


```
Db 1926 lntHrSerLeuSerLysProSerSer-----SerIlePheGlnAsnTrpP 1942
QY 3698 CCTGCCACAGTACTCTCCAGAGATGACGACTTCTGGCTCCTGAAGAGCTGCAG 3757
Db 1942 IolysProPro---LeuProPro-----LeuProThrglySerGlyValSers 1957
QY 3758 ACCTGGCTGTGGCTCGCGCCAGGACTTCAACGGCTCAAGAGAGATGCAGCTCCA 3817
Db 1957 erSerThrAlaAlaAlaProglyAlaThrSerProSer-----AlaSerSerAlaSerA 1975
QY 3818 GCAGCTCAGTCACCTGCACCTGGGGCTCATGCTTCTGACTTCTGACTTCTCCTCT 3877
Db 1975 laSerThrArgHis---LeuGlnGlyValGluPheGluMetArgProProLeuLeuArgA 1994
QY 3878 TCGCTCCC---CCTTCA-----AACCCTGCTCCCACTT 3907
Db 1994 rgAlaProSerProSerLeuLeuProAlaSerAspHisLysValSerProAlaProArg- 2013
QY 3908 TGTGAGAGCCAGCCCTGTATGCCAACA---CCTGTGAGCCAGGAGACAGAGCTGTGAG 3964
Db 2014 ProSerSerLeuProIleLeuProSerGlyProIleTyrProGlyLeuPheAspIleArg 2033
QY 3965 CCTCTGCCCTTTCCTCGACCGCTGGCGGTGATGGATCAGCCCTGTCTCCTCCCA 4024
Db 2034 SerSerProThrGlyGlyAlaGlySerThrAspProPheAlaProValPheValPro 2053
QY 4025 CCTCCCAAGGTCTACCGAGTGGGAGGAGTACAGTAGGCCCTGTCTCTCTCTGTTTC 4084
Db 2054 ProHisProGlyIle-----SerGlyGlyLeuGlyAlaLeuSer----- 2067
QY 4085 TACAGGAAGTCATGCTCGAGGAGTGTCAAAGTGTTCAGTTGTTGTCAGAGCGCTCATG 4144
Db 2068 -----GlyAlaSerArgSerLeuSerProThrArgLeu----- 2078
QY 4145 GCCTCCTGCTTCTTGCTTACCACTTGCCAGTGCACCCAGCCCTCAGGTGGCACATC 4204
Db 2079 -----LeuSerLeuProProAspLysProPheGlyAlaLysProLeu----- 2092
QY 4205 TGGAGGCGAGGGTTGAGGGGCCACCACACATGCTTCTGGGGTGAA----- 4255
Db 2093 -----GlyPheTrpThrLysPheAspVal 2100
QY 4256 GCCCTTTGGTGGCCCACTCTCCTTGGATGGGTGTTGCTCCCTTATCCCCAAATCACTCT 4315
Db 2101 AlaAspTrpLeu-----GluTrpLeuGlyLeuSerGluHisArgAlaGlnPheLeu 2117
QY 4316 ATACATCCAATTCCAGGAACAAACATGTCGCAATT 4351
Db 2118 AspHisGluIleAspGlySerHisLeuProAlaLeu 2129
```

Search completed: January 27, 2003, 16:20:13
Job time : 530.364 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:43:12 ; Search time 214.014 Seconds
(without alignments)
6334.602 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aactcgagtgaggcctggc.....ccttgtaagtctctca 5087

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+, n2p.model, -DEV=xlp
-Q/cgn2_1/USFPOO/US09931704/runat 27012003 154125 3585/app_query.fasta_1.7189
-DB=A Geneseq 101002 -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704.CGN 1 1 222 @runat 27012003 154125 3585 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseq 101002:.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1446	15.3	321	22	ABBI1896 Human cardiostroph
2	1446	15.3	321	22	AAW79399 Human protein SEQ
3	1112	11.8	260	22	AAW78415 Human protein SEQ
4	910.5	9.7	253	21	AAW25831 Human protein sequ
5	896	9.5	215	21	AAW19586 Human interleukin-
6	896	9.5	225	19	AAW29715 Human neurotrophic
7	896	9.5	225	19	AAW56141 Amino acid sequenc
8	896	9.5	225	22	AAW94466 Human cardiostroph
9	896	9.5	225	21	AAW78133 Human NNT-1 protei
10	896	9.5	225	22	AAW63543 Amino acid sequenc
11	896	9.5	225	23	AAW78176 Human novel neurot
12	885	9.4	164	22	ABBA0317 Peptide #7823 enco
13	885	9.4	164	22	ABBA24716 Protein #6715 enco
14	885	9.4	164	22	AAW61118 Human brain expres
15	885	9.4	164	22	AAW73827 Human bone marrow
16	885	9.4	164	22	AAW20115 Peptide #6549 enco
17	885	9.4	164	22	AAW34012 Peptide #8049 enco
18	885	9.4	164	23	ABGA3716 Human peptide enco
19	884	9.4	223	22	AAE00828 Human cardiostroph
20	863	9.1	215	21	AAW19587 Mouse interleukin-
21	863	9.1	225	19	AAW29716 Mouse neurotrophic
22	863	9.1	225	19	AAW56142 Amino acid sequenc
23	863	9.1	225	21	AAW87814 Murine NNT-1 prote
24	863	9.1	225	23	AAW78177 Mouse novel neurot
25	496	5.3	94	20	AAW12768 Human 5' EST secre
26	396.5	4.2	763	18	AAW31852 Mycobacterium tube
27	382.5	4.1	1047	22	ABW70874 Drosophila melanog
28	378.5	4.0	2284	22	ABW71434 Human apolipoprote
29	376.5	4.0	1008	22	ABW11527 Novel human diagno
30	374	4.0	1013	22	ABW08112 Novel human diagno
31	366	3.9	4561	22	ABG30203 Novel human diagno
32	366	3.9	9222	22	ABG21064 Novel human diagno
33	364	3.9	2639	22	ABG15016 Novel human diagno
34	351.5	3.7	2087	22	ABG15156 Amino acid sequenc
35	341.5	3.6	2703	22	ABG63299 Drosophila melanog
36	339.5	3.6	1212	20	AAW87503 Human N-methyl-D-a
c 37	334	3.6	1078	16	AAW1704 Collagen alpha 1 (
38	334	3.5	2091	21	AAW12000 Collagen type III
c 39	330.5	3.5	1078	21	AAW96125 Collagen type III-
c 40	330.5	3.5	1078	23	ABW09628 Collagen type III-
c 41	330.5	3.5	1078	23	ABW09628 Amino acid sequenc
c 42	330.5	3.5	1078	23	AAE16478 Human collagen alp
c 43	330.5	3.5	2971	21	AAW41231 Human ORF995
c 44	329.5	3.5	1469	22	AAW15191 Novel human diagno
c 45	327.5	3.5	1466	22	ABW50291 Collagen type III

ALIGNMENTS

RESULT 1

ABBI1896

ID ABBI1896 standard; peptide; 321 AA.

XX ABBI1896;

AC ABBI1896;

XX

DT 11-JAN-2002 (first entry)

XX

DE Human cardiostrophin-like cytokine homologue, SEQ ID NO:2366.

XX

KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;

cell culture; drug screening; gene therapy; antiinflammatory;
antischismatic; antiarthritic; haemostatic; antiarteriosclerotic;
cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
antifungal; vulnery; antiulcer.

XX Homo sapiens.

XX WO200157188-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03800.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

XX N-PSDB; ABA09140.

XX Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -

XX Claim 20; Page 273; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
invention also relates to vectors and recombinant host cells comprising a
nucleotide of the invention, methods of producing the novel polypeptides,
antibodies against the polypeptides, methods of detecting the nucleotides
or polypeptides in a sample, and methods of identifying compounds which
bind to polypeptides of the invention. Although novel, many of the
polypeptides of the invention have homology to known proteins, thereby
giving an insight into their probable biological activities, and hence
potential therapeutic applications. The polypeptides of the invention may
have various activities, including cytokine, cell proliferation or cell
differentiation activities; stem cell growth factor activity;
haematopoiesis regulatory activity; tissue growth activity;
immunomodulatory activity; activin- or inhibin-related activities;
chemotactic or chemokinetic activities; haemostatic, thrombotic or
thrombolytic activities; receptor or ligand activities; or may be
involved in oncogenesis, cancer cell proliferation or metastasis.
Depending on their biological activities, polypeptides and nucleotides of
the invention are useful for preventing, treating or ameliorating medical
conditions, e.g., by protein or gene therapy. Such conditions include
cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
proliferative retinopathy, atherosclerosis, coronary heart disease,
arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
vascular growth. Polypeptides involved with tissue regeneration and
repair (or nucleic acids encoding them) may be used to promote wound
healing (e.g., of burns, incisions and ulcers), while those with
immunomodulatory activities may be used in the treatment of viral,
bacterial and fungal infections in addition to immune disorders.
Polypeptides with growth factor activity may be used in cell cultures to
promote cell growth. For example, such polypeptides may be used to
manipulate stem cells in culture to give rise to neuroepithelial cells
that can be used to augment or replace cells damaged by illness,
autoimmune disease or accidental damage. The polypeptides and nucleotides
may also be used in the diagnosis of the above conditions, and in drug
screening techniques. The present sequence represents a novel human
polypeptide of the invention.

XX Sequence 321 AA;

Alignment, Scores:

Pred. No.: 1,2e-96 Length: 321

Score: 1446.00 Matches: 319

Percent Similarity: 34.48% Conservative: 2

Best Local Similarity: 34.26% Mismatches: 0
Query Match: 15.33% Indels: 610
DB: 22 Gaps: 1

US-09-931-704-3 (1-5087) x ABB11896 (1-321)

Qy 1064 TTGGCACACACAGTGGCACACATGCCAAAGACTCTCTCAGCTGACACACAGATCCATT 1123
Db 1 ValAlaHisThrGlnTrpHisThrCysGlnArgLeuSerGlnLeuThrHisArgSerIle 20

Qy 1124 CTCAGTATCTACTGATAGACATCATCGTGGCCAAAGTCTCTCATCTCAAAACATACAT 1183
Db 21 LeuLysTyrLeuLeuIleAspThrHisAlaCysGlnValLeuIleLeuLysHisThrHis 40

Qy 1184 GCCTCTCTTCTCTCCCGTCTTCCAGGAGTGTTCCTCCCTCCTCATCCCTCTGCTCC 1243
Db 41 AlaSerLeuSerLeuProSerCysGlnGluCysPheProSerSerIleProSerAlaSer 60

Qy 1244 CATCTGTGTCCACCTCTCACCCCCACCCAGCCCAAGTGGGGACAGACACTGTAGGGG 1303
Db 61 HisMetValSerHisProHisProProProSerProArgTyrGlnThrProGluGly 80

Qy 1304 CTGCAGCTGTCTCCCGTGTGGGCCCGCGCGCGCTCATGTCTTCGTCTCATCTGCCCC 1363
Db 81 LeuProAlaAlaSerProCysGlyProGlyProArgSerCysPheSerSerIleLeuPro 100

Qy 1364 ACAGGGACTCGTGGGGATGTAGCGTGCCTGTGCACGGTCTCTGGCACCTCCCTGCA 1423
Db 101 ThrGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAla 120

Qy 1424 GTGCCAGCTCTCAATGCGACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTAT 1483
Db 121 ValProAlaLeuAenArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyr 140

Qy 1484 GACCTACCCGCTACTCGGACACCAACTCCGACGCTGGCTGGGACCTATGTGAGTATC 1543
Db 141 AspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThrTyr 157

Qy 1544 CAGCGTAGGAATCTGGGAGTTGGGGAGGAGTGAGGAGTTGGGGAAAGACAGTCCTAACCG 1603
Db 157 ----- 157

Qy 1604 TGGAGGTTCTGGTAAATGATGGGGTGAGAGGGGCTCTTTGGCTCCACACAGTCCCTT 1663
Db 157 ----- 157

Qy 1664 GTCTGGTCTATCTCTGCTTCCCTCTTAGTGGCCCCCCCCACTTCCCATCCCTGGCC 1723
Db 157 ----- 157

Qy 1724 CCAGGACTAGGCATGTGGGCAGGCTCGCACCGCTTGGGCCATTTGCCCATCGGCTGC 1783
Db 157 ----- 157

Qy 1784 CAGCCCAGCCGCCGCTCTCCCTCTGGGGGCCCGGGGAAGTCTCTCTGTTTACACCGTGT 1843
Db 157 ----- 157

Qy 1844 GTGGTGTCTTTCGCGGGCGGGGTGGGTGGGGACAGAGGGGGCCCCACCTCCCATGCCT 1903
Db 157 ----- 157

Qy 1904 GCGTTCAGCTCGGCTCTGCCCCCAGACCTGGGGGCCCTGTCTCTGGACCCAGGGGCT 1963
Db 157 ----- 157

Qy 1964 CCCTTCGCTGTGCCTCTCCCATCTTAGTGGGGCTCTTAGGGGGGTCTATGGGGGAAGGGG 2023
Db 157 ----- 157

Qy 2024 ACTGTAGGAACCCAGGCAGTAGTGGCAGGGGGTTTAGGTGTGGATGGAGTTATGCTG 2083
Db 157 ----- 157

PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejрман T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK52532.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 237; 6221pp; English.
 CC
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AMW8323-AMW8302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AMW8020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 321 AA;

Alignment Scores:
 Pred. No.: 1,2e-96 Length: 321
 Score: 1446.00 Matches: 319
 Percent Similarity: 34.48% Conservative: 2
 Best Local Similarity: 34.26% Mismatches: 0
 Query Match: 15.33% Indels: 610
 DB: 22 Gaps: 1

US-09-931-704-3 (1-5087) x AAM79399 (1-321)

QY 1064 TTGGCACACACAGTGGGCACATGCGCAAGACTCTCTCAGCTGACACAGATCCATT 1123
 Db 1 ValAlaHisThrGlnTrpHisThrCysGlnArgLeuSerGlnLeuThrHisArgSerIle 20
 QY 1124 CTCAGTATCTACTGATAGACACTCATGCTGCGCAAGTCTCTCTCTCAACATACACAT 1183
 Db 21 LeuLeuTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
 QY 1184 GCCTCTCTTCTCTCCGCTTCCAGGAGTGTTCCTCCCTCTCTCTCTCTCTCTCTCTCTCT 1243
 Db 41 AlaSerLeuSerLeuProSerCysGlnGluCysPheProSerSerIleProSerAlaSer 60
 QY 1244 CATCTGCTCTCCACCTCACCCCTCCAGGAGTGTTCCTCCCTCTCTCTCTCTCTCTCTCTCT 1303
 Db 61 HisMetValSerHisProHisProProSerProArgTrpGlyGlnTrpGlyGlnTrpGly 80
 QY 1304 CTGCCAGCTGCTTCCCGCTGTGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1363
 Db 81 LeuProAlaAlaSerProCysGlyProGlyProArgSerCysPheSerSerIleLeuPro 100
 QY 1364 ACAGGGGACTCTGGGGGAGTGTAGCGTCTGTGACCGTGTCTGTGGCACCTCCCTGCA 1423
 Db 101 ThrGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAla 120
 QY 1424 GTGCCAGCTCTCAATCGCACAGGGGACCCAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1483

Db 121 ValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyr 140
 QY 1484 GACCTCACCCGCTACTCGAGGACCACTCCCGAGCTTGGCTGGGACCTATGTGAGTATC 1543
 Db 141 AspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThrTyr 157
 QY 1544 CAGCGTAGGAATCTGGGAGTGTGGGAGGAGTGGAGAGTTGGGGAAGACAGTCTTAACCG 1603
 Db 157 157
 QY 1604 TGGAGGGTTCTGGTAAATGATGGGTGAGAGGGGCTCTTTGGCTCCCACTCCCT 1663
 Db 157 157
 QY 1664 GTCTGGTCTATCTCTCGCCCTTCCCTCTTAGGTGGCCCCCCTTCCCACTCCCTGGCC 1723
 Db 157 157
 QY 1724 CCAGGACTAGGCATGTGGGAGGCTCGCACCCGCTTGGCCCATGTGCCCATGCTGCTGC 1783
 Db 157 157
 QY 1784 CAGCCACGCCCGCCCTCCCTCTGGGGCCGGGAAAGTCTCTGTGTTTACACCGTGT 1843
 Db 157 157
 QY 1844 GTGGTGTCTCTTGGCGGGCGGGTGGGTGGGACAGAGGGGCCCACTCCCATGCCT 1903
 Db 157 157
 QY 1904 GCGTTCCAGCTCGCTCTGCCCCCAGACTGGGGGCCCTGCTCTCTGGACCCAGGGGCT 1963
 Db 157 157
 QY 1964 CCCTTCCGCTGCTCTCTCCCATCTAGTGGGCTCTCTAGGGGGTCTATGGGGGAAGGG 2023
 Db 157 157
 QY 2024 ACTGTAGGNAACCCAGGACGATAGTGGCAGGGGTTTAGGGTGTGGATGGAGTTATGCTG 2083
 Db 157 157
 QY 2084 TAAGGATTTGGGGGTGGTCCAGAGGTGTTTTCAGAGAGCCAGGAGAGAGAGAGGGT 2143
 Db 157 157
 QY 2144 GGAGAGCCGAGGACCACTGGGGAACCGGCCCTCTTCCCGTGTCTCTCTTCCACATCC 2203
 Db 157 157
 QY 2204 CAGACCCTACTCTGAGGCCAGGAAAGAAAGGAAAGAGTGGCGGGGAGCTGCTCC 2263
 Db 157 157
 QY 2264 AGCCCCAGGATPACACAGGAAATTAGTTGTCTGTCTGTGTGTGTGTGTGTGTGTGTGT 2323
 Db 157 157
 QY 2324 CCCTGGGCGCTTGGCTATCCAGGCTCTCCCTTGTCTCTCTCTCTCTCTCTCTCTCTCT 2383
 Db 157 157
 QY 2384 CATCTCCCTCATCTTTCCTGGGCCCGCCAGCGCTCCCGAGGGTGGAAAGGCTCT 2443
 Db 157 157
 QY 2444 GCCCTCTTCCCTATPACCATGCTGTCTTCCATAGCCTTCTCTCTCTCTCTCTCTCTCTCT 2503
 Db 157 157
 QY 2504 GCCTCCATTTCT 2563

Db 157 ----- 157
Qy 2564 TAGTGAGTACCGTCTCTCCCGAGCCCTCAGCTGGTGGCGCTGGTGTGTGAGCGCAA 2623
Db 157 ----- 157
Qy 2624 ATGGGGCTCTGTTTCCAATGGGCCACTCTCATCTCTCTCTTCTTCTGTCGAGAAAACC 2683
Db 157 ----- 157
Qy 2684 TTTGGCTTCACTCCACTGCGCTCTCTAGTTCCCGACCCCTTTTCTCTCTGCTGGCTTCCCTG 2743
Db 157 ----- 157
Qy 2744 CCAATTTTCCAGGAGTGGTCTACACCTCTGCTCTCCACTTCTCTCCACCCACTCAC 2803
Db 157 ----- 157
Qy 2804 TTCTTAACCCCTGCAATCTGGCTTCCAGGCCCCAGCAATGTTCTCTCAAGGTGCTCA 2863
Db 157 ----- 157
Qy 2864 GGCACCTCTTCCAGCGCGACAGTGTITGAAGGCTCATCTCTCTGCTCTCTGTTT 2923
Db 157 ----- 157
Qy 2924 GCAGCCACACTGCTGAGCGCTGCTGCTTTCGAACTCTCTTCTCTGCTCTCTGCACTC 2983
Db 157 ----- 157
Qy 2984 TCCTGGGCACTTCTACCTCTCCAGCTCTCCAGGCTCTCTCTCTCTCTCTCTGCTG 3043
Db 157 ----- 157
Qy 3044 CCACAGCGGCACTCTCCAGGTTTGGCCACCCAGCCCAATCAGCAGCTCTCTCTGAGC 3103
Db 157 ----- 157
Qy 3104 GTCTTGTGGTCT 3163
Db 157 ----- 157
Qy 3164 CCACTGTTCACTGTCACTGCATACAAATGATATCTTATTTGAAATACTCAGGAGG 3223
Db 157 ----- 157
Qy 3224 CCATGAACAAAGCCTAGCTAGTGGAGACAGGGCCAGTCTCAGGGGACACAAAAATAGA 3283
Db 157 ----- 157
Qy 3284 AACTTTGGAGCAGGTATCTCTTGTGTGGTGGAGCCAGCGCTCTGCGCTCTCTCTCTCC 3343
Db 157 ----- 157
Qy 3344 ATCACCTCTCTTTTCACTGTAACACTGAGGCGCCCTTTTCAACAGCACTTC 3403
Db 158 -----LeuAsnTyrLeuGlyProProPheAsnGluProAspPhe 170
Qy 3404 AACCTCCCGCTGGGGCAGAGCTCTGCCAGGCGCACTGTTCACTGGAGGTG 3463
Db 171 AsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGluValTrp 190
Qy 3464 CGAAGCCTCAATGACAACTGGCGCTGACCCAGAACTACGAGCGCTTACAGCCACTTCTG 3523
Db 191 ArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeu 210
Qy 3524 TGTACTTCGGTGGCTCAACCGTACGCTGAGTGGCACTGCTGAGTGGCGCGCACTGGCC 3583
Db 211 CysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAla 230
Qy 3584 CACTTCTGCACCACTCCAGGCGCTCTGCGGAGCAATTCGGGCGCTCATGCACTGCTG 3643
Db 231 HisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeu 250

Qy 3644 GGCTACCCACTGCCCGCCAGCCGCTGCTGGAGTGAACCCACTTGACTCTGCTGGCCCTGCC 3703
Db 251 GlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGlyProAla 270
Qy 3704 CACAGTGACTTCTCCAGAGATGAGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3763
Db 271 HisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeuGlnThrTrp 290
Qy 3764 CTGTGGCGCTCGCCCAAGCACTTCAACCGGCTCAAGAAAGATGAGCTCCAGCAGCT 3823
Db 291 LeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnProProAlaAla 310
Qy 3824 GCAGTCACTCTGCACTGGGGCTCATGCTTC 3856
Db 311 AlaValThrLeuHisGlyAlaHisGlyPhe 321
RESULT 3
AAW78415
ID AAW78415 standard; Protein; 260 AA.
XX
AC AAW78415;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 1077.
XX
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Zhang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51548.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3306; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

Db 96 ----- 96
 Qy 3227 TGAACAAAGACCTAGCATGAGACAGCGCCAGTGTGAGGGACACAAAATAGAAAC 3286
 Db 96 ----- 96
 Qy 3287 TTTGGAGACGATATCTCTTGGTGGTGGAGCAGCGGCTGTGCCCTCTCCCTTCCCATC 3346
 Db 96 ----- 96
 Qy 3347 ACCCTCTCTTTACACAGTGAACCTACTGCGCCCGCCCTTCAACAGCAGCAGCTTCAAC 3406
 Db 97 -----LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsn 110
 Qy 3407 CTTCCCGCTGGGGCAGAGACTGTGCCCGGGCCACTGTTGACTTGGAGGTGGGCA 3466
 Db 111 ProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArg 130
 Qy 3467 AGCTCAATGACAACTCGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGT 3526
 Db 131 SerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCys 150
 Qy 3527 TACTTGCTGGCTCAACGCTCAGCTGCCACTGCTGAGCTGCGCGCAGCTGCGCCAC 3586
 Db 151 TyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHis 170
 Qy 3587 TTCTGCACAGCTCCAGGGCTGTGGGCGCATTTGGGGGTCTATGGCAGCTCTGGGC 3646
 Db 171 PheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGly 190
 Qy 3647 TACCCACTGCCCGCTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCAC 3706
 Db 191 TyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGlyProAlaHis 210
 Qy 3707 AGTGACTTCTCCAGAGTGAACACTTCTGGCTGCTGAGAGAGCTGAGACCTGGCTG 3766
 Db 211 SerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeu 230
 Qy 3767 TGGCGCTCGGCCAAGACTTCAACCGCTCAAGAGAGATCAGCTCCAGCAGCTGCA 3826
 Db 231 TrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnProProAlaAlaAla 250
 Qy 3827 GTACCTCTGCACCTGGGGCTCATGGCTTC 3856
 Db 251 ValThrLeuHisLeuGlyAlaHisGlyPhe 260
 RESULT 4
 AAM25831
 ID AAM25831 standard; Protein; 253 AA.
 AC AAM25831;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:1346.
 XX
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnerary; antilucer; osteopathic; eczema;
 KW dermatological; allergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopoenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 XX

OS Homo sapiens.
 FN WO200153455-A2.
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000WO-US35017.
 XX
 PR 23-DEC-1999; 99US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PT Tang YT, Liu C, Drmanac RT;
 PT WPI: 2001-457603/49.
 DR N-PSDB; AAH99772.
 DR
 XX
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 Claim 20; Page 278; 1217pp; English.
 PS
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 AAM25963. The proteins can have activities based on the tissues and
 cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis of
 CC disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopoenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 XX
 SQ Sequence 253 AA;
 Alignment Scores:
 Pred. No.: 9.64e-58 Length: 253
 Score: 910.50 Matches: 231
 Percent Similarity: 26.84% Conservative: 2
 Best Local Similarity: 26.61% Mismatches: 14
 Query Match: 9.65% Indels: 621
 DB: 4
 US-09-931-704-3 (1-5087) x AAM25831 (1-253)
 Qy 1253 TCCACCTCACCCTCCACCCAGCCCAAGTGGGGGACACACCTGAGGGGCTGCCAGCT 1312
 Db 7 AlaArgProSerThrProProAlaSerGlyArgGlyAlaAlaPro---GlyArgPro--- 24
 Qy 1313 GCTTCCCGCTGTGGGCCCGCGCGCTCATGCTTCTCTGCTCCATCTCTGCCACAGGGGAC 1372
 Db 25 -----GlyProSerProMetAsp-----LeuArgAlaGlyAsp 35
 Qy 1373 TCGTGGGGAGATTAGGTGCTGTGCGAGGTGCTGTGCGACCTCCCTCGCTGCCAGCT 1432
 Db 36 SerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAla 55
 Qy 1433 CTCATCCGACAGGGGACCCAGGGCTGGCCCTCCATCCAGAAACCTATGACCTCAC 1492
 Db 56 LeuAsnArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThr 75


```

Db 186 LeuProGlnProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAsp 205
Qy 3713 TTCCTCCAGAGATGGACGACTTGGCTGCTGAAGAGCTGCAGACCTGGCTGGCGC 3772
Db 206 PheLeuGlnLysMetAspPheTrpLeuLysGluLeuGlnThrTrpLeuTrpArg 225
Qy 3773 TCGGCCAAGGACTTCAACCGCTCAAGAAGAGATGCAGCTCCAGAGCTGCAGTCAAC 3832
Db 226 SerAlaLysAspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThr 245
Qy 3833 CTGCACCTGGGGCTCATGGCTTC 3856
Db 246 LeuHisLeuGlyAlaHisGlyPhe 253

RESULT 5
ID AAB19586 standard; Protein; 215 AA.
XX
AC AAB19586;
XX
DT 22-JAN-2001 (first entry)
XX
XX Human interleukin-B60 (IL-B60).
XX
KW Interleukin-B60; IL-B60; human; cytokine; cytokine-like factor-1;
KW haematopoietic; inflammation; antiinflammatory; autoimmune disease;
KW therapy.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..17
XX FT Peptide /label= Signal_peptide
XX FT Protein 18..215
XX FT /label= Mature-protein
XX
XX WO200053631-A1.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06182.
XX
XX 11-MAR-1999; 99US-0267901.
XX
XX (SCHE ) SCHERING CORP.
XX
XX Oppmann B, Timans JC, Kastelein RA, Bazan JF;
XX
XX WPI; 2000-587426/55.
XX
XX N-PSDB; AAA88546.
XX
XX Cytokine-like factor 1 (CLF-1) and interleukin (IL)-B60 complexes, and for
XX polypeptides, and nucleic acids, useful in research, diagnosis and for
XX treating inflammatory and autoimmune disorders -
XX
XX Claim 1; Page 15-16; 97pp; English.
XX
XX The present sequence is that of human interleukin-B60 (IL-B60), a
XX novel, small soluble cytokine-like protein that exhibits structural
XX motifs characteristic of a member of the long-chain cytokines, and
XX which shows homology to granulocyte colony stimulating factor and
XX interleukin-6. IL-60B may have either stimulatory or inhibitory
XX effects on haematopoietic cells, including e.g. lymphoid cells,
XX such as T-cells, B-cells, natural killer cells, macrophages,
XX dendritic cells, haematopoietic progenitors, etc. Methods are
XX provided for modulating the physiology or development of a cell or
XX tissue culture cells by contacting the cell with an agonist or
XX antagonist of IL-B60 or an agonist of antagonist of a complex of
XX mature IL-B60 and its partner, cytokine-like factor-1 (CLF-1, see
XX AAB19588). The IL-B60/CLF-1 cytokine serves as a key physiological
XX factor in motor neuron development and regeneration. IL-60B, its
XX agonists and antagonists may be used to treat inflammatory or

```

```

CC autoimmune disorders and also for drug screening.
XX
XX Sequence 215 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1.03e-56 Length: 215
XX Score: 896.00 Matches: 170
XX Percent Similarity: 90.72% Conservative: 6
XX Best Local Similarity: 87.63% Mismatches: 6
XX Query Match: 9.50% Indels: 12
XX DB: 2
XX

US-09-931-704-3 (1-5087) x AAB19586 (1-215)
Qy 3311 GGTGAGCCAGCGGCTCTGCCTCC-----TCCTTCCCATCACC----- 3349
Db 22 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 41
Qy 3350 -----CTCTCTTTTTCACAGCTGAACCTACCTGGGCGCCCTTTCAACGAG 3394
Db 42 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGlu 61
Qy 3395 CCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTCTTGTACTTG 3454
Db 62 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 81
Qy 3455 GAGGTGTGGCAAGCTCAATGACAACTGGGGCTGACCCAGAACTACGAGGCTACAGC 3514
Db 82 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 101
Qy 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGAGCTGCCACTGTGAGTGGCCGC 3574
Db 102 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArg 121
Qy 3575 AGCTTGGCCCACTTCTGCACAGCTCCAGGGCTGCTGGGCGAGCATTCGCGGCGTCATG 3634
Db 122 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 141
Qy 3635 GCAGCTCTGGGCTACCCACTGCCCGCCGCTGCTGGGACTGAACCCACTTGGACTCTCT 3694
Db 142 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpPro 161
Qy 3695 GGCCCTGCCCACTGAGTCTTCTCCAGAGATGGAGCAGCTTCTGGCTGCTGAAGGAGCTG 3754
Db 162 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLysGluLeu 181
Qy 3755 CAGACTGGCTGTGGGCTCGGCCAAGGACTTCAACCGCTCAAGAAGAGATGCAGCCT 3814
Db 182 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 201
Qy 3815 CCAGCAGCTGCAGTCACTGCCTGCACCTGGGGGCTCATGGCTTC 3856
Db 202 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 215

RESULT 6
AAW29715
ID AAW29715 standard; Protein; 225 AA.
XX
XX AAW29715;
XX
XX 09-NOV-1998 (first entry)
XX
XX Human neurotrophic factor NNT-1.
XX
XX NNT-1; neurotrophic factor; human; antiinflammatory; adjuvant;
XX Alzheimer's disease; Parkinson's disease; Huntington's disease;
XX amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
XX peripheral neuropathy; dystrophy; neural retina degeneration;
XX common variable immunodeficiency; CVID; selective IgA deficiency;
XX hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
XX therapy.
XX
XX OS Homo sapiens.

```


Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
 QY 3515 CACCTTCTGTGTTACTTGGCTGAGCTCAACCGTCAAGGCTGCACCTGCTGAGCTGCGCGC 3574
 Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131
 QY 3575 AGCTGCGCCACTTCTGCACAGCTCCAGGCGCTGCTGGGCGCATTTGCGGCGCTCATG 3634
 Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
 QY 3635 GCAGCTTGGCTACCCACTGCGCCAGCGCTGCTGGGCTGAGCTGAGCTGAGCTGCT 3694
 Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
 QY 3695 GGCCTCTCCACAGTCTCTCCAGAGATGAGCACTTCTGGCTGCTGAAGGAGCTG 3754
 Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191
 QY 3755 CAGACTGCTGGCTGCGCTCGGCGCAAGCACTTCAACCGGCTCAAGAAGAAGATGAGCT 3814
 Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
 QY 3815 CCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3856
 Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 9
 AAY87813
 ID AAY87813 standard; Protein; 225 AA.
 AC AAY87813;
 DT 24-AUG-2000 (first entry)
 XX Human NNT-1 protein.
 DE NNT-1; human; neurotrophic factor; neurotrophic; neuroprotective; treatment;
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
 KW retinopathy; immune disorder; hematopoietic disorder.
 XX Homo sapiens.
 XX US6054294-A.
 PN 25-APR-2000.
 PD 12-DEC-1997; 97US-0988819.
 PF 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 PA Chang M;
 PI WPI; 2000-338492/29.
 DR N-PSDB; AAA39481.
 XX New nucleic acids encoding neurotrophic factors useful for stimulating
 PT growth of motor or sympathetic neurons for treating neuron cell damage
 PT
 XX Claim 1c; Fig 3; 42pp; English.
 XX This invention describes a novel nucleic acid molecule (I) encoding a
 CC novel neurotrophic factor (NNT-1) (II) which has neurotrophic,
 CC neuroprotective, anticonvulsant, antiparkinsonian, antidiabetic and
 CC ophthalmological activity. (I) is useful for producing NNT-1
 CC polypeptides which are useful for treating patients in whom various
 CC cells of the central, autonomic, or peripheral nervous system have

CC degenerated and/or have been damaged by congenital disease, trauma,
 CC mechanical damage, surgery, stroke, ischemia, infection, metabolic
 CC disease, nutritional deficiency, malignancy and/or toxic agents. NNT-1
 CC proteins are used to treat diseases like Alzheimer's, Parkinson's,
 CC amyotrophic lateral sclerosis, Charcot-Marie-Tooth syndrome, Huntington's
 CC disease, peripheral neuropathy induced by diabetes or other metabolic
 CC disorders, and/or dystrophies or degeneration of the neural retina such
 CC as retinitis pigmentosa, drug-induced retinopathies, stationary forms of
 CC night blindness, progressive cone-rod degeneration, immune disorders and
 CC hematopoietic disorders. (I) is effective in treating neurological
 CC conditions and promotes neuron regeneration. Neural functions are
 CC effectively restored in patients suffering from various neurological
 CC disorders. This sequence represents the human NNT-1 protein described in
 XX the method of the invention.

SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 1.05e-56 Length: 225
 Score: 896.00 Matches: 170
 Percent Similarity: 90.72% Conservative: 6
 Best Local Similarity: 87.63% Mismatches: 6
 Query Match: 9.50% Indels: 12
 DB: 21 Gaps: 2

US-09-931-704-3 (1-5087) x AAY87813 (1-225)

QY 3311 GGTGAGCCAGCGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
 Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
 QY 3350 -----CTCTCTCTTTTTCACAGCTGAACTACTGCGGCGGCTTTTCAACGAG 3394
 Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
 QY 3395 CAGACTTCAACCTTCCCGCTCGGCGGCGAGAGACTCTGCCAGGCGCCTGTTGACTTG 3454
 Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
 QY 3455 GAGGTGTGGCAAGCTCAATCACAACTGCGGCTGACCCAGCAACTACGAGGCTACAGC 3514
 Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
 QY 3515 CACCTTCTGTGTTACTTGGCTGCTCAACCGTCAAGGCTGCCACTGCTGAGCTGCGCGC 3574
 Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131
 QY 3575 AGCTGCGCCACTTCTGCACAGCTCCAGGCGCTGCTGGGCGCATTTGCGGCGCTCATG 3634
 Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
 QY 3635 GCAGCTCTGGGCTACCCACTGCGCGCGCTGCTGGGAGCTGAACCCACTTGGACTCCT 3694
 Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
 QY 3695 GGCCTCTCCACAGTCTCTCCAGAGATGAGCACTTCTGGCTGCTGAAGGAGCTG 3754
 Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191
 QY 3755 CAGACTTGGCTGCGCTCGGCGCAAGCACTTCAACCGGCTCAAGAAGAAGATGAGCT 3814
 Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
 QY 3815 CCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3856
 Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 10
 AAG63543
 ID AAG63543 standard; Protein; 225 AA.
 XX AAG63543;
 AC AAG63543;
 XX

US-09-931-704-3 (1-5087) x ABB40317 (1-164)

QY 3365 CTGAATACCTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCTGGGGCA 3424

Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20

QY 3425 GAGACTGCCCCAGGCGCACTGTGACTGGAGGTGGGAGAGCTCAATGCAAACTG 3484

Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40

QY 3485 CGGCTGACCCAGAACTACAGAGCCACAGCCACTTCTGTGTTACTTGGTGGCTCCAC 3544

Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60

QY 3545 CGTACAGGCTGCCACTGTGAGCTGCGCGCAGCTTGGCCCTTCTGACACGCTCCAG 3604

Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeuGln 80

QY 3605 GGCCTGTGGGCGACATTTGGGCGTCAATGGAGCTTGGGCTACCCACTGCCCCAGCGG 3664

Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGlnPro 100

QY 3665 CTGCTGGGAGCTGAACCCACTTGGACTCTGGCCCTGCGCCCTCAAGTCTCTCCAGAG 3724

Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120

QY 3725 ATGACGACGCTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGTGGGCTACCCACTGCCCCAGCGG 3784

Db 121 MetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLysAsp 140

QY 3785 TTCAACCGGCTCAAGAGAGATGACAGCTCCAGAGCTGAGTCCACCTGACCTGGGG 3844

Db 141 PheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeuGly 160

QY 3845 GCTCATGGCTTC 3856

Db 161 AlaHisGlyPhe 164

RESULT 13

ABB24716

ID ABB24716 standard; Protein; 164 AA.

XX AC ABB24716;

DT 23-JAN-2002 (first entry)

DE Protein #6715 encoded by probe for measuring heart cell gene expression.

XX KW Human; gene expression; heart; microarray; vascular system;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease.

OS Homo sapiens.

XX WO200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

PS Claim 15; SEQ ID No 26486; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, the human heart and vascular system e.g. cardiovascular disease, diagnosing, grading, staging, monitoring and prognosing diseases of the hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 164 AA;

QY

Alignment Scores:

Pred. No.: 5,91e-56 Length: 164

Score: 885.00 Matches: 163

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 99.39% Mismatches: 0

Query Match: 9.38% Indels: 0

DB: 22 Gaps: 0

US-09-931-704-3 (1-5087) x ABB24716 (1-164)

QY 3365 CTGAATACCTGGGCCCCCTTTCAACAGAGCCAGACTTCAACCCCTCCCGCTGGGGCA 3424

Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20

QY 3425 GAGACTGCCCCAGGCGCACTGTGACTGGAGGTGGGAGGCTCAATGCAAACTG 3484

Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40

QY 3485 CGGCTGACCCAGAACTACAGAGCCCTACAGCCACTTCTGTGTTACTTGGTGGCTCCAC 3544

Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60

QY 3545 CGTACAGGCTGCCACTGTGAGCTGCGCGCAGCTTGGCCCTTCTGACACGCTCCAG 3604

Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeuGln 80

QY 3605 GGCCTGTGGGCGACATTTGGGCGTCAATGGAGCTTGGGCTACCCACTGCCCCAGCGG 3664

Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGlnPro 100

QY 3665 CTGCTGGGAGCTGAACCCACTTGGACTCTGGCCCTGCGCCCTCAAGTCTCTCCAGAG 3724

Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120

QY 3725 ATGACGACGCTTCTGGCTGTGAAGAGCTGCAGACCTGGCTGTGGGCTACCCACTGCCCCAGCGG 3784

Db 121 MetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLysAsp 140

QY 3785 TTCAACCGGCTCAAGAGAGATGACAGCTCCAGAGCTGAGTCCACCTGACCTGGGG 3844

Db 141 PheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeuGly 160

QY 3845 GCTCATGGCTTC 3856

Db 161 AlaHisGlyPhe 164

RESULT 14

AAM61118

ID AAM61118 standard; Protein; 164 AA.

XX AC AAM61118;

XX

DB: 22 Gaps: 0

US-09-931-704-3 (1-5087) x RAN73827 (1-164)

QY 3365 CTGAACTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCGCCCTGGGGGCA 3424
|||||
Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20
|||||

QY 3425 GAGACTCTGCCCGGCCACTCTTGACTTGGAGGTGTGGCGAAGCCTCAATGACAAACTG 3484
:::|||||
Db 21 LysThrLeuProArgAlaThrValaspLeuGluValTyrArgSerLeuAsnAspLysLeu 40
|||||

QY 3485 CGGCTGACCCAGAACTACGAGCCCTACAGCCACTTCTGTGTACTTGTGGTGCCTCAAC 3544
|||||
Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60
|||||

QY 3545 CGTCAGGCTGCCACTGCTGAGCTGGCGCCGAGCTGGCCACTTCTGCACCCAGCCTCCAG 3604
|||||
Db 61 ArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeuGln 80
|||||

QY 3605 GGCCTGTCTGGCGAGCATTCGCGGCGTCAATGGCAGCTCTGGGCTACCCACTGCCCCAGCCG 3664
|||||
Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGlnPro 100
|||||

QY 3665 CTGCTGGGACTGAACCCACTTGGACTCTGGCCCTGCCCCACAGTCACTTCTCCAGAAG 3724
|||||
Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120
|||||

QY 3725 ATGGACGACTTCTGGCTGCTGAAGGCTGCAGACCTGGCTGTGGCTGGCCCAAGGAC 3784
|||||
Db 121 MetAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLysAsp 140
|||||

QY 3785 TTCAACCCGCTCAAGAAGAGATGCAGCCTCCAGCAGCTGCAGTCACTCACCTGCACCTGGGG 3844
|||||
Db 141 PheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeuGly 150
|||||

QY 3845 GCTCATGCTTC 3856
|||||
Db 161 AlaHisGlyPhe 164
|||||

Search completed: January 27, 2003, 15:59:49
Job time : 235.014 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 16:20:23 ; Search time 36.4279 Seconds

(without alignments)
5635.701 Million cell updates/sec

Title: US-09-931-704-3

Perfect score: 9432

Sequence: 1 aacctgcgagtgccctggc.....ccttgctaagtctctcca 5087

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 244452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09931704/runat_27012003_154129_3760/app_query.fasta_1.7189
-DB=Published Applications AA -QFMT=fastan -SURFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=51 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09931704 @CGN 1 1 51 @runat_27012003_154129_3760
-NCPU=6 -ICPU=3 -NO_XLPV -NO_MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	896	9.5	225	US-09-931-704-2
2	885	9.4	164	US-09-864-761-40014
3	863	9.1	225	US-09-931-704-5
4	281.5	3.0	1806	US-09-919-497-56

Sequence 2, Appli	1274	2.8	267	9	US-10-020-215-2
Sequence 10, Appli	714	2.8	259	10	US-09-861-597-10
Sequence 187, App	595	2.6	239.5	9	US-09-854-133-187
Sequence 187, App	595	2.6	239.5	10	US-09-738-973-187
Sequence 1068, Ap	5179	2.5	239.5	9	US-10-025-380-1068
Sequence 1068, Ap	5179	2.5	239.5	10	US-09-922-217-1068
Sequence 2, Appli	5179	2.5	239.5	11	US-09-833-263-1068
Sequence 2, Appli	233	2.5	233	12	US-09-835-232-2
Sequence 56, Appli	1806	2.5	233	13	US-09-919-497-56
Sequence 3, Appli	230.5	2.4	230.5	14	US-09-735-367B-3
Sequence 2, Appli	2063	2.4	226	15	US-09-735-367B-2
Sequence 5, Appli	1690	2.4	223.5	16	US-09-788-043C-5
Sequence 2, Appli	877	2.4	222.5	17	US-10-041-770-2
Sequence 425, App	4019	2.3	218.5	9	US-09-854-133-425
Sequence 435, App	4019	2.3	218.5	10	US-09-738-973-425
Sequence 24, Appli	507	2.3	216.5	9	US-10-078-547-24
Sequence 108, App	1367	2.3	216	10	US-09-801-368-108
Sequence 2, Appli	802	2.3	215	11	US-09-823-240-2
Sequence 2, Appli	503	2.3	214	9	US-10-078-547-2
Sequence 7, Appli	529	2.3	213.5	10	US-09-861-597-2
Sequence 53, Appli	1317	2.3	212.5	9	US-09-963-896-7
Sequence 4, Appli	646	2.2	204.5	10	US-09-964-899-53
Sequence 17, Appli	606	2.2	203.5	9	US-09-861-597-4
Sequence 15, Appli	645	2.1	198.5	10	US-09-964-899-17
Sequence 6, Appli	593	2.1	198	9	US-09-964-899-15
Sequence 12, Appli	1234	2.1	197.5	10	US-09-861-597-6
Sequence 10, Appli	1056	2.1	196.5	9	US-09-854-173A-12
Sequence 1, Appli	651	2.1	195.5	10	US-10-161-510-10
Sequence 34, Appli	1940	2.1	195	9	US-09-861-597-1
Sequence 50, Appli	1134	2.1	193.5	10	US-10-016-283-34
Sequence 8, Appli	606	2.1	192.5	9	US-10-001-873-50
Sequence 47, Appli	826	2.1	192.5	10	US-09-861-597-8
Sequence 108, App	638	2.0	191	9	US-09-894-998-47
Sequence 2, Appli	1655	2.0	190.5	10	US-10-001-887-108
Sequence 5, Appli	2630	2.0	190.5	9	US-09-858-664A-2
Sequence 17, Appli	7968	2.0	190.5	10	US-10-077-130-2
Sequence 2, Appli	731	2.0	190	9	US-10-077-130-5
Sequence 17, Appli	503	2.0	188.5	10	US-10-086-464-17
Sequence 2, Appli	1618	2.0	188.5	9	US-10-078-547-2
Sequence 1, Appli	2442	2.0	188.5	10	US-09-963-875-1
Sequence 10, Appli	2442	2.0	188.5	12	US-10-109-886-10

ALIGNMENTS

RESULT 1

US-09-931-704-2
; Sequence 2, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating Ige-Related Disease Using Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-931-704-2

Alignment Scores:	4.84e-43	Length:	225
Pred. No.:	896.00	Matches:	170
Score:	90.72%	Conservative:	6
Best Local Similarity:	87.63%	Mismatches:	6
Query Match:	9.50%	Indels:	12
DB:	10	Gaps:	2


```
US-09-931-704-3 (1-5087) x US-09-931-704-2 (1-225)
QY 3311 GGTGAGCCAGCGCTCTCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTCTTTTTCACAGCTGAACCTACTCTGGGCCCCCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTCGGGGCGAGAGACTCTGCCAGGGCCACTGTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGCGAGGCTCAATGACAACTGCGGCTGACCCAGAACTACGAGGCTACAGC 3514
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAAGCTGCGCTGCACTGCTGAGCTGCGCCG 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGGGCCACTTCTGCACAGCTCCAGGGCCCTGCTGGGAGAGCATTCGGGCGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 CGAGCTGGGCTACCCAGTCCCGCCAGCGCTGCTGGGAGAGCATTCGGGCGCTCATG 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
QY 3695 GGCCCTGCCACAGTACTCTCTCAGAAAGATGGAGCACTTCTGGCTGCTGAAGGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLysGluLeu 191
QY 3755 CAGACTGGCTGTGGCTGCGCCAGGACTTCAACCGCTCAAGAGAGATGCGAGCTT 3814
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
QY 3815 CCAGCAGCTGCAGTCACTGACCTGGGGGCTCATGGCTTC 3856
Db 212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 2
US-09-864-761-40014
; Sequence 40014, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmika-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST HUMAN HIT: A1752561.1, EVALUE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 1,96e-42 Length: 164
Score: 885.00 Matches: 163
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.39% Mismatches: 0
Query Match: 9.38% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-3 (1-5087) x US-09-864-761-40014 (1-164)
QY 3365 CTGAACCTACCTGGGCCCCCTTTTCAACGAGCAGACTTCAACCTTCCCGCTGGGGCA 3424
Db 1 LeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20
QY 3425 GAGACTCTGCCAGGGCCACTGTGTGACTTGGAGGTGTGGCGAAGCTCAATGACAACTG 3484
Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40
QY 3485 CGGCTGACCCAGAACTACGAGGCTACAGCCACTTCTGTGTACTTGGTGGCTTCAAC 3544
Db 41 ArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeuAsn 60
QY 3545 CGTCAGCTGCCTGCTGAGCTGCGCGCAGCTGCGCCACTTCTGCACCAAGCTTCCAG 3604
Db 61 ArgGlnAlaAlaThrAlaGluLeuArgSerSerLeuAlaHisPheCysThrSerLeuGln 80
QY 3605 GGCCTGCTGGGCGAGCATTCGGGGCGTTCATGGCAGCTCTGGGCTACCCACTGCCCCAGCCG 3664
Db 81 GlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGlnPro 100
QY 3665 CTGCTGGGACTGAACCCACTTGGACTCTGCGCTGCCCGCCACAGTACTTCTCCAGAG 3724
Db 101 LeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGlnLys 120
```


Qy	1738	C-----ATGCTAGTCTCTGGGGCCAGGGATGGGAACTGGGG-----	1700
Db	1441	lyLeuIysGlyAspProGlySerLys--GlyGluLysGlyHisProGlyLeuIleGlyL	1460
Qy	1701	-----GGGCACCTAAGAGGGAAGGCGAGAGATACACAG-----	1666
Db	1460	eulleGlyProGlyGluGlnGlyGluLysGlyAspArgGlyLeuProGlyThrGlnG	1480
Qy	1665	-----ACAGGGGACTGTGGGAGCCAAAGAGCCCTCTCTCACCC	1626
Db	1480	lySerProGlyAlaLysGlyAspGlyGlyIleProGlyProAlaGlyProLeuGlyProp	1500
Qy	1625	CATCATTTACAGAACCTCCACGGTTAGACATGCTTTTCCCACTCTCACTCTCTCC	1566
Db	1500	roGly-----	1501
Qy	1565	CAACTCCCAGATTCTACGCTGGATACTCACATAGTCTCCAGCCAAAGCTCGGAGTTGGT	1506
Db	1502	-----ProProGlyLeu-----	1505
Qy	1505	GCTCCAGGTAGCGGTAGGTATAGTGTTCGTGGATGGAGGGCCAGGCGCTCTGGTCCC	1446
Db	1506	-----ProGlyProGlnGlyP	1511
Qy	1445	CTGTGCGATTGAGAGTGCACCTGCAGGGAGGTGCAGAGACACCGTCACAGGCACGCTA	1386
Db	1511	roLysGlyAsnLysGlySerThr-----	1518
Qy	1385	ACATCCCCACAGTCCCTCTGTGGCAGGATGCAGAGAGCATGACGCGGC---CCGG	1329
Db	1519	-----GlyProAlaGly-----GlnLysGlyAspSerGlyLeuProG	1511
Qy	1328	GOCCACAGGGGAAGCAGCTGCGAGCCCTCAGGTCTGTCTCCACCTTGGGTGGGTG	1269
Db	1531	lyProProGly--ProProGlyProProGlyGluValleGlnProLeuProIleLeuS	1550
Qy	1268	GGGGTGAGGGTGGACACACCATGGAGGCGAGAGGGATG-----GAGGAGG	1211
Db	1550	er-----SerLysIysThrArgHisThrGluGlyMetGlnAlaAspAlaAspAspA	1568
Qy	1220	GGAAACACTCTGGCAGACGGGAGAGAA	1192
Db	1568	enlleLeuAspTyrSerAspGlyMetGlu	1577
RESULT 5			
US-10-020-215-2			
; Sequence 2, Application US/10020215			
; Publication No. US20030008347A1			
; GENERAL INFORMATION:			
; APPLICANT: PLOWMAN, GREGORY			
; APPLICANT: PELES, EIOR			
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF ALP RELATED DISORDERS			
; FILE REFERENCE: 038602/1290			
; CURRENT APPLICATION NUMBER: US/10/020,215			
; CURRENT FILING DATE: 2001-12-18			
; PRIOR APPLICATION NUMBER: 09/095,443			
; PRIOR FILING DATE: 1998-06-10			
; PRIOR APPLICATION NUMBER: 60/049,477			
; PRIOR FILING DATE: 1997-06-11			
; NUMBER OF SEQ ID NOS: 10			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 2			
; LENGTH: 1274			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Descriptive of Artificial Sequence: ALP			
; OTHER INFORMATION: polypeptide sequence			
US-10-020-215-2			
Alignment Scores:			
Pred. No. : 5,98e-08 Length: 1274			
Score: 267.00 Matches: 288			

[illegible]

547	CysPheProValProProProGlnProLeuProThrProTyrThrTyrProAlaGlyAla	566
Qy	AAGCTAGGGTACAACACAGCAGCAGTGCAGCTGCAGTCCCTAGGCCTGGCCCCGT	1038
Db	LyS-----GlnProIle	570
Qy	CCCTCCATGTACACACATATACATGTTGGCACACACACAGTGGCACACATGCCAAGACT	1098
Db	ProAlaGlnHis--HisPheSerGlyIleProThrGlyPheProAlaProArg---	588
Qy	CTCTCAGCTGCACACACAGATCCATCTTCTCAAGTATCTACTGATAGACACTCATCGTGCCA	1158
Db	IleGlyProGln-----Pro	593
Qy	AGTCCTCATCTCAACATACATACATCGCTCTCTTCTCTCCCGTCTTGCCA---GGAGTG	1215
Db	GlnProHisProGlnProHisProSerGlnAlaPheGlyProGlnProProGlnGlnPro	613
Qy	TTCCTCCCTCTCCATCCCTCTGCCTCCCATCTGGTGTCCACCTCACCCTCCACCCACCCAG	1275
Db	LeuProLeuGlnHisProHisLeuPheProProGlnAlaProGlyLeuLeuProProGln	633
Qy	CCCAAGTGGGACACACACCTGAGGGGTGCCAGCTGCTCCCGTGTGGGCCCGGCC	1335
Db	SerProTyrPro-----	637
Qy	GGCTCATGCTTCTCGTCCATCTGCTCCACAGGGGACTCGTGGGGATGTTAGCGTCCT	1395
Db	TyrAlaProGlnProGlyValLeuGlyGlnProProProPro	651
Qy	GTGCACGGTGCTCTGGCACTCCCTGCAGTGCAGCTCTCAATGCACAGGGGACCCAGG	1455
Db	LeuHisThrGlnLeuTyrPro-----GlyProAla	661
Qy	GCCTGGCCCTC-----CATCCAGAAACCTATGACCTCAACCCGTACTCTGGAGCACA	1509
Db	GlnAspProLeuProAlaHisSerGlyAlaLeu---ProPheProSerProGlyProPro	680
Qy	ACTCGCAGCTTGGCTGGGACCTATGTAGTATCCAGCGTAGGAATCTGGAGTGGGGA	1569
Db	GlnPro-----ProHisProProLeuAlaTyrGlyProAlaPro	693
Qy	GGAGTGAGGAGTTGGGAAAGACAGTC-----CTAACCTGGAGGTTCTGTGTAATGA	1623
Db	SerThrArgProMetGlyProGlnAlaAlaProLeuThrIleArgGlyProSer-----	711
Qy	TGGGTGAGGGGCTCTTGGCTCCACAGTCCCCCTGTCTGTCTATCTCTCTGCC	1683
Db	SerAlaGlyGlnSerThrProSerProHisLeuValPro	724
Qy	TTCCTCTTAGTGGCCCCCCTTCCCATCTCCCTGGCCACGAGGACTAGGCATGTGGC	1743
Db	SerProAla-----ProSerProGlyPro-----	732
Qy	AGGCCTCGCACCCGCTTGGCCCATTCGCCACTGGCTGCCAGCCAGCGCCCGCTCC	1803
Db	GlyProValProProArgProProAla-AlaGluProProPr	746
Qy	CCCTGGGGCGGGGAAGTCTCTCTGTTACCGT-----GTTGT	1845
Db	CysLeuArgArgGlyAlaAlaAlaAspLeuLe	758
Qy	GGTGTCTTGGCGGGGGGTTCGGTGGGACAGAGGGGCCACCTCCCATGCTGC	1905
Db	UserSerSerProGluSerGlnHisGlyGlyThrGlnSerProGlyGlyGlnProLe	778
Qy	GTTCAG-----CTCGCTCTCGCCCGACCTGGGGCCCTGTGCTCTG	1950
Db	LeuGlnProThrIleValAspAlaAlaGluGlyArgArgProGlnAlaLeuArgLeuI	798
Qy	GACCACGGGCGCT-----CCCTTCGCTGT-----CCTCTCCATCTCTAGCTGGCCT	1998
Db	eGluArgAspProTyrGluHisProGluArgLeuArgGlnLeuGlnGlnGluAl	818

OTHER INFORMATION: Description of Artificial Sequence:protein
US-09-861-597-10

Alignment Scores:

Fragment Scores:	
Pred. No.:	1.63e-07
Score:	259.00
Percent Similarity:	28.12%
Best Local Similarity:	22.93%
Query Match:	2.77%
DB:	10
Length:	714
Matches:	230
Conservative:	52
Mismatches:	332
Indels:	389
Gaps:	50

US-09-931-704-3 (1-5087) x US-09-861-597-10 (1-714)

Qy	3391	CCAGCCGCTCCAGGAAG-----GGCCAGAGGCTCAGAGTTCTGTCT	3395
Db	2	ProSerGlyProGlySerAlaAlaAlaAlaAlaGlyPro-GlyGlnGln-----G1	19
Qy	3349	TCCTGGCTCAACAGGTGTTGGCATACAGGCTGGCTCTCACAAGTGGGAGCAGGGTTG	3890
Db	19	yProGlyGlyTyrglyProGly---Gln-GlnGlyProGlyGlyTyrglyProGlyGlnG	38
Qy	3889	AAGGGGAGCGAAGAGGAAAGGTCAAGAGTCAGAAGCCATGAGCCCCCAGGTGCAGGGT	3830
Db	38	lnGly-----ProSerGlyProGly-44	44
Qy	3829	GACTGCAGCTGTGGAGGCTGCATCTTCTCTGTGACCGGTTGAAGTCTTGTGCCGAGCG	3770
Db	44	-----44	44
Qy	3769	CCACAGCCAGGTCTGCAGCTCTCTTACAGACCCAGAAGTCGTCCA-----	3726
Db	45	-----SerAlaAlaAlaAlaAlaAlaAlaGlyProGlyGlyTyrglyProG	61
Qy	3725	--TCCTTCGGAGAGTCACTGTGGCAGGGCCAGAGAGTCCAAGTGGTTCAGTCCCGAG	3668
Db	61	ylGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyProSerGlyProGlySerA	81
Qy	3667	CAGCGCTGGGCGAGTGGGTAGCCCGACAGCTGCCATGACGCCCGCAATGCTGCCACGAC	3608
Db	81	laAlaAlaAlaAlaAlaAla-----AlaG	90
Qy	3607	GCCCTGGAGCTGGTCGAAAGTGGGCCAGAGCTGGCGGCGAGCTCAGCAGTGGCAGGCTG	3548
Db	90	lyProGlyGlyTyr-----GlyProGly-----GlnGlnGlyProG	102
Qy	3547	ACGGTTGAGCCACGCAAGTAACACAGAGGTGGCTGTAGGCCCTCGTAGTCTCTGGG---	3492
Db	102	lyGlyTyrglyProGlyGlnGlnGlyProGlyGlyTyrglyProGlyGlnGlnGlyProS	122
Qy	3491	-----TCAGCCGCAAGTTTGTTCATTGAGGCTTCGCCACACCT-----	3456
Db	122	erGlyProGlySerAlaAlaAlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyT	142
Qy	3455	-----CCAAGTCAACAGTGGCCCTGGCAGAGTCTGTGCCCCAGCGGGAGGTTGAA	3401
Db	142	yrGlyProGlyGlnGlnGlyProGlyGlyTyrgly-----ProGlyGlnGlnGlyProS	160
Qy	3400	GTCTGGCTCGTTCAAGGGGGGGCCAGGTAGTTTCAGCTGTGAAGAGAGAGGTTGATGGG	3341
Db	160	er-----GlyProGlySerAlaAlaAlaAlaAlaAlaAlaGly	173
Qy	3340	GAAGGAGGA-----GGGCAGACCGGTGGCTCACCACCAAGGAGATACCTGTCTCC	3290
Db	174	ProGlyGlyTyrglyProGlyGlnGln-----182	182
Qy	3289	AAAGTTTCTATTTTTTGTGTGCCCTGCACACTGGCCCTGTCTCCTATGCTAGGCTCTTTGT	3230
Db	182	-----182	182
Qy	3229	TCATGGCCCTCCCTGAGTTTTTCCAATAAGGATATCATTTGTATGCAAGTTCAGAGTGAAG	3170
Db	182	-----182	182

Qy	3169	CAGTGGCGGTGGTGAGCTCTCCATGGAGAGCGCTAGAAAAAAGGAGGAGGAGGACGACA	3111
Db	183	-----GlyProGlyGlyProGlyProGly	190
Qy	3109	CAAGACGCTCAGGAAGGACGTGCTGATTGGCTGGGTGGGCAAACCTTGGGAGAGTCCCG	3050
Db	191	GlnGlnGlyProSerGlyProGlySerAlaAlaAlaAlaAlaAlaAlaAlaGlyPro	210
Qy	3049	CTGTGGGGCCAGGACAGAGAGGAAGAGAGCCCTGGAGGAGCTGGAGAGGTAGAAAGTGGC	2990
Db	211	---GlyGlyTyrglyProglyGlnGlnGlyProGlyGlyTyrglyProglyGlnGlnGly	229
Qy	2989	CCAGAGAGTGCAGAGACCAAGAGAGAGGTTCGAGAAAGGCAGCAGC-----GCTCAG	2936
Db	230	ProGly-----GlyTyrglyProgly---GlnGlnGlyProSerGlyProGlySer	245
Qy	2935	CAGTGTGCTGCAAAACAGACAGCAGCAAGGANGATGAGCCTTCAAAACACTGTGGGCTTG	2876
Db	246	AlaAlaAlaAlaAlaGlyProglyGlnGlnGlnGlyProglyGlyTyrglyProglyGln	2855
Qy	2875	CAAGCA---GGTGCCTGACGACCTTGGAGAGAACCATTCCTGGGGCTCGAAGCCAGATT	2819
Db	266	GlnGlyProGlyGlyTyrglyProglyGlnGlnGlyProSerGlyProglySerAlaAla	285
Qy	2818	GCAGGG-----GGTTAAGAAGTGAGTGGTGGAGAGAAAGTGGAGGC	2777
Db	286	AlaAlaAlaAlaAlaGlyProglyGlyTyrglyProglyGlnGlnGlyProglyGly	305
Qy	2776	AGAGGTGTAGCCTCTTGGAGAAATTTGGCAGGGAAGCCAGGAGAGAAAAAGGGT	2717
Db	306	TyrGly-----ProglyGln-----GlnGlyProSerGlyProglySerAla	319
Qy	2716	CGGGAACTAGAGAGGGCAGTGAGTCAAGCAAAAGGTTTTCTGCACAAGCAAGAGAGA	2657
Db	320	AlaAlaAlaAlaAlaAlaAlaGlyProglyGlyTyrglyProglyGlnGlnGlyPro	339
Qy	2656	GATGAGAGTGGCCCATTTGGAACAGAGCCCATTTCCGCTGCACACCCAGCCCAACCA	2597
Db	340	GlyGlyTyrglyPro-----GlyGlnGlnGlyProgly	350
Qy	2596	GCTGAGGGCTGGGAGAGACGGGTACTCACTAACCTCGAAAGAGGGTTTCAGCTGAT	2537
Db	351	GlyTyrglyProglyGlnGlnGlyProSer-----GlyProglySerAlaAla	366
Qy	2536	AGGACAGGGTTGCAGAGAGAAATGA-----GGCAGTCTCATGAGTAGG	2489
Db	367	AlaAlaAlaAlaGlyProglyGlnGlnGlyProglyGlyTyrglyProglyGlnGlnGly	386
Qy	2488	ACAGGAGGAAGGCTATGGAAGACAGCATGGTATAGGAAGAGGGCAGAGCCCTTCCAAC	2429
Db	387	ProglyGly-----TyrGlyProglyGlnGlnGly	396
Qy	2428	CCTCGGGG-----ACGCGCTGGGGC	2408
Db	397	ProSerGlyProglySerAlaAlaAlaAlaAlaAlaAlaGlyProglyGlyTyrgly	416
Qy	2407	CCAGGGAAGGATGAGGAGATGTATACTGGAAAGAAGGGAGAGCAAGGGGAGAGG	2348
Db	417	Progly-----GlnGlnGly-----	421
Qy	2347	CCTGGGATAGCAAGGGCCCGAGGGAGGTTCACAGCTGCACAGCACAGAGACAACATA	2288
Db	422	Progly---GlyTyrglyProglyGlnGlnGlyProSerGly-----	434
Qy	2287	ATTTCTCGGTGATCTCTGGGCTGGAGCGAGCTCCCCCGCACCTTCTTCCCTTTCTT	2228
Db	435	-----ProglySerAlaAlaAlaAlaAlaAla-----	444
Qy	2227	TCCCTGGCTCAGAGTAGGCTCTGGATCTGGAAGAGCAACCGGGAAGAGGGCGCGT	2168
Db	445	AlaAlaGlyProgly-----GlyTyrglyProglyGlnGlnGlyProglyGlyTyr	461
Qy	2167	TCCCATGTGTGCTGGCTCTCCACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2108

Db 199 -----ArgGlySerArgMetGlyArgGlnHisGluAla---AlaAlaThrAlaAla 214
Qy 2503 AGTCTCATGATGAGCAGGAGGCGTATGGAAGCAGCATGTTATGGAAGAGGCG 2444
Db 215 ThrAlaAlaThrAlaThrGlyGlyThrAlaGluGluAlaGlyAlaSerAlaProGluSer 234
Qy 2443 AGAGCCCTTTTCCAACTCGGGGAGCGGTGGGGCCAGGAAAGGATGAGGGAGATG 2384
Db 235 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnGlyArgArg 254
Qy 2383 TATACTGGGAAGAGG-----GAGNAGCAAGGGGAGAGCCTGGG 2342
Db 255 HisGlyThrGlnArgArgArgGlyProProGlnAlaArgGluGluGlyProArgAla 274
Qy 2341 -----ATAGGCAAGGCG---CCAGGGGAGGTTCCACACCTGACACAGCAGAGAC 2294
Db 275 ThrThrIleGluGlyLeuGlyThrProSerGlyGluGlnArgAlaAspGln----- 291
Qy 2293 AAATAATTCTCGGTGTATCCT-----GGGGCTGGAGCCAGCTCCCGCCGACCC 2243
Db 292 -----SerGlnAlaLeuProAlaLeuAlaGlyAlaAlaAlaHis--AlaHisA 308
Qy 2242 TCTTCCCTTTCTTCCCTCGGTCCAGAGTAGGTTCTGGATGTGGAGAGGAGACAGG 2183
Db 308 laileProGlyAlaGlyProAlaAlaAlaProValGlyArgGlyArgGlyGlyT 328
Qy 2182 GAAGAGGGCGGTTCCCATGCTCGGTGCTCGGTCTCCCAACCTCCTTCTCTCTCT 2123
Db 328 rArgGlyGlyArg----- 332
Qy 2122 GGGCTCTCTGAACACCTCTGGACCCCAATCTTACAGCATACCTCCATCCACAC 2063
Db 332 ----- 332
Qy 2062 CCTAAACCCCTGCCACTACTGCTGGTTCCTACAGTCCCTTCCCGCATGACCCCC 2003
Db 333 -----A 333
Qy 2002 TAGGAGGCCAGCTAGGATGGGAGAGCAGACGGAAGGGAGGCCCTGGTCCAGAGCAG 1943
Db 333 rgGlySerAlaGlyAlaGly--GlyGlyGlyArg-----GlyGlyArgGly- 348
Qy 1942 CAGGGCCCGGCTCTGGGGCAGAGCGAGCTGGAACGAGGATGGAGGTGGGGCCCC 1883
Db 349 ArgGly--ArgGlyGlyArgGly--GlyGlyGlyAlaGlyArgGlyGlyGlyAla-- 366
Qy 1882 TCTGTCCCAACCCCGCCCGCAGAGCAGACACACACACGCTGTAAACAGAGAG 1823
Db 367 ----AlaGlyProArgGlyAlaSerSerProGlyAlaArgArgGlyGluGlnArg 385
Qy 1822 ACTTCCCGCG---CCCCAGGGGAGGCGG-----CGGCT 1790
Db 385 gArgGlyArgGlyProProAlaAlaGlyAlaAlaGlnValSerAlaArgGlyArgAl 405
Qy 1789 GGGCTGGCAGCCAGTGGGCAATGGGCCAAGCGGGTGGCGAGGCTGCCACATGCCCTAG 1730
Db 405 aArgGlyGlnArgAlaGlyGluGluAlaGlnAspGlyLeuLeuProArgGlyArgAsp 425
Qy 1729 TCTTGGGGCAGGATGGGAAGTGGG-----GGGCCACTAGAG 1688
Db 425 gLeuProLeuArgProGlyAlaAlaAlaAsnGlnArgAlaGluArgProGlyProArg-- 444
Qy 1687 GGAAGGCGAGGATAGACACAGAGCGGCGACTGTTGGGAGCCAAAGAGCCCTCTCAC 1628
Db 445 -----GlyGly----- 446
Qy 1627 CCCATCATTTACAGAACCTCCAGCTTAGGACTGTCTTCCCACTCTCTACTCTCTC 1568
Db 447 -----HisGlyProValAsnAlaSerSerAlaProAspThrSe 459
Qy 1567 CCCAACTCCAGATTCTACGCTGGATCTACTACATAGTCTCCAGCCAGCTGGAGTTG 1508
Db 459 rPro-----ProArgHisProArgArgtr 467

Qy 1507 GTGCTCCAGGTAGCGGTGAGGTATAGTCTTCTGGATGAG----- 1465
Db 467 pValSerGlnGlnArgGlnArg-----LeuTrpArgGlnPheArgValGlyGly 484
Qy 1464 -----GGCCAGGCGCTGGTCCCT-----GTCGGATTGAG 1433
Db 484 yPheProProProProSerArgProProAlaValLeuLeuProLeuLeuArgLeuAl 504
Qy 1432 A---GCTGGCACTCAGGGAGGTGCCAGACACCGTCACAGCAGCGTAACTCCCCCA 1376
Db 504 aCysAlaGlyAspProGlyAlaThrArgProGlyProArgArgProAlaArgProAr 524
Qy 1375 C----- 1375
Db 524 gGlyGluLeuIleProArgArgProAspProAlaAlaProSerGluGluGlyLeuArgMe 544
Qy 1374 -GAGTCCCTGTG-----GGCAGATGAGAGAGCATGAGCGCGCGGCCCCACCA 1322
Db 544 tGluSerSerValAspAspGlyAlaThrAlaThrThrAlaAlaSerGlyGluAl 564
Qy 1321 CGGGGAAGCAGCTGCAGCCCTCAGGTGTCTGTCCCACTTGGGCTGGGTGGGGGTG 1262
Db 564 aProGluAlaGlyProSerProSerHis-Ser---ProThrMetCysGlnThrGlyGlyP 583
Qy 1261 AGGTGGGACACCA 1248
Db 583 rGlyProProPro 587

RESULT 8

US-09-738-973-187
; Sequence 187, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Segrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-187

Alignment Scores:
Pred. No.: 1.98e-06 Length: 595
Score: 239.50 Matches: 184
Percent Similarity: 33.29% Conservative: 59
Best Local Similarity: 25.21% Mismatches: 240
Query Match: 2.57% Indels: 247
DB: 10 Gaps: 39

US-09-931-704-3 (1-5087) x US-09-738-973-187 (1-595)

Qy 3130 AAAGCGAGGAG-----GGAGACGACAGAGCGCTCAGGAGGACGTGTGTAT 3083
Db 8 GluGlyGlyGlyGluTrpGlyProGlyValProArgGluArgGluSerAlaGlyGlu 27

RESULT 9
US-10-021

; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Alignment Scores:

Pred. No.:	2,168-06	Length:	5179
Score:	239.50	Matches:	285
Percent Similarity:	27.58%	Conservative:	118
Best Local Similarity:	19.51%	Mismatches:	500
Query Match:	2.54%	Indels:	560
DB:	9	Gaps:	59

US-09-931-704-3 (1-5087) x US-10-025-380-1068 (1-5179)

Qy	1129	GTATCTACTGATAGACACTCATCGTCCCAAGTCCTCATCTCAAAACATACATGCCTC	1188
Db	1030	ValSerThrAsnProGluProCysLeuAsnProHisArgSerTrpAlaGluLys	1049
Qy	1189	TCTTCTCTCCGCTCTGCCAGGAGTGTTCCTCTCTCCATCCCTCTGCTCCCACTCT	1248
Db	1050	GlnCysSerIleLeuLysSerValPheSerIleCysHis	1064
Qy	1249	GGTGTCACCCCTCACCCCCACCCAGCCCAAGTG	1296
Db	1065	LysValAspProLysProPheTrpGluAlaCysValHisAspSerCysAspThr	1084
Qy	1297	TGAGGGGCTGCCAGCTGCTTC	1317
Db	1085	GlyGlyAspCysGluCysPheCysSerAlaValAlaSerTyrAlaGlnGluCysThrLys	1104
Qy	1318	-----CCGCTGGGGCCCGCGCTCATGCTTCTC-----	1353
Db	1105	GluGlyAlaCysValPheTrpArgThrProAspLeuCysProIlePheCysAspTyrTyr	1124
Qy	1354	CATCTGGCCAC	1365
Db	1125	AsnProHisGluCysGluTrpHisTyrGluProCysGlyAsnArgSerPheGluThr	1144
Qy	1366	-----AGGGGACTCGTGGGGATGTTAGCGTGCCTGTGACCGGTGCTCGCA	1413
Db	1145	CysArgThrIleAsnGlyIleHisSerAsnIleSerValSerTyrLeuGluGlyCysTyr	1164
Qy	1414	CCTCCCTGCGAGTCCAGCTCTCAATCCACAGGGGACCCAGGGCTCGGCCCTCCATCCA	1473
Db	1165	ProAfgCys-----ProllysAspArgProIleTyrGlu	1175

Qy	1474	GAATAACCTA-----	1491
Db	1176	GluAspLeuLysLysCysValThrAlaAspLysCysGlyCysTyrValGluAspThrHis	1195
Qy	1492	CCGCTACTGGAGCACCACCTCCGCGAGCTTGGTGGGACCTATGTGAGTATCCAGCGTAG	1551
Db	1196	TyrProProGlyAlaSerValProThr-----GluGluThrCysLysSerCysValCys	1213
Qy	1552	GAATCTGGGAGTTGGGGAGGAGTCAGGAGTTGGGAAAGACAGATCTCTA-----ACCCTG	1605
Db	1214	ThrAsnSerSerGlnValValCysArgProGluGluGlyLysIleLeuAsnGlnThrGln	1233
Qy	1606	GAGGTTCTCGTAAATGATGG-----GGTAGAGAGGGGCTCTTTGGCTCCACCCAG	1656
Db	1234	AspGlyAlaPheCysTyrTrpGluIleCysGlyProAsnGlyThrValGluLysHisPhe	1253
Qy	1657	TCCCTCTGTCTGTCTATCTCTCTGCC-----TTCCTCTTAGGTGGCCCC	1701
Db	1254	AsnIleCysSerIleThrThrArgProSerThrLeuThrThrPheThrThrIleThrLeu	1273
Qy	1702	CCCCACTTCCCATCTCCCTGGCCCCCAGGACTAGGCATGTGGCGAGGCCT-----CGCAC	1755
Db	1274	ProThrThrProThrSerPheThrThrThrThrThrThrThrThrThrThrThrThr	1293
Qy	1756	CGCTTGGCCCCATTGCC-----CACTGGCTGCCACCCAGCCGCG	1794
Db	1294	ValLeuSerThrThrProLysLeuCysCysLeuTrpSerAspTrpIleAsnGluAspHis	1313
Qy	1795	CCGCTCTCCCTCGGGCGGGGAGTCTCTCTGTGTACACCGTGTGTGTGTGTCTCT	1854
Db	1314	ProSerGlySerAspAspGluProPheAspGlyValCysGly-----	1331
Qy	1855	TGCGCGGGCGGGTGGGTGGGGACAGAGGGCCCCACCTCCCATGCTCGCTCCAGCT	1914
Db	1332	-----AlaProGluAspIleGluCysArgSerVal	1341
Qy	1915	CGCTCTGCCCCAGACCTGGGGCCCTGCTCTGTGGACCCAGGGGCTCTCTTCCGTCT	1974
Db	1342	Lys-----AspProHisLeuSerLeuGluGln	1350
Qy	1975	GCCTCTCCCATCTAGTGGGCCCTCTAGGGGGTCTAGGGGAAGGGGAC-----	2025
Db	1351	-----HisGlyGlnLysValGlnCysArgSerVal	1361
Qy	2026	-----TGTAGGAACCCAGCAGTAGTGGCAGGGGTTTAGGGTGTGATGGAGGT	2076
Db	1362	GlyPheIleCysLysAsnGluAspGlnPheGlyAsnGlyProPheGlyLeuCysTyrAsp	1381
Qy	2077	TATGCTGTAAAGATT-----TGGGGGTGTCTCAGAGGTGTTCAGAGAGCCAGGAG	2127
Db	1382	TyrLysIleArgValAsnCysCysTrp--ProMetAspLysCysIleThrThrPro---	1399
Qy	2128	AGAAGGAAGAGGGTTGGAGAGCGGAGCCAGCCATCGGGGAACCGGCCCTCTTCCCGTG	2187
Db	1400	-----SerProProThrThrThrProSerProProThrThrThrThr	1414
Qy	2188	TTCCTCTTCCACATCCACAGACCTACTCTGGAGCCAGGAAAGAAAGGAAGAGTGG	2247
Db	1414	hrThrLeuPro-----	1417
Qy	2248	CGGGGAGCTGGCTCCAGCCCCCAGGATACCCGAGGAATTAGTTTGTCTCTGTGTGTGT	2307
Db	1418	--ProThrThrThrProSerProThrThr-----T	1428
Qy	2308	CAGCGTGTGAACCTCCCTCGGCCCTTGCCTATCCAGG-----C	2348
Db	1428	hrThrThrThrProProThrThrThrProSerProProIleThrThrThrThrThrP	1448
Qy	2349	CTCTCCCT	2408
Db	1448	roLeuProThrThrThrProSerProIleSerThrThrThrThrThrProProThrThr	1468

QY 2408 CCCAGCGCTCCCGGAGGCTGGAAAGGCTGCGCCTCTCCCTATACCATGCTGC 2468
Db 1468 hrThrProSerProPro-----ThrThrThrProSerPro----- 1479
QY 2469 TTCATAGCTTCCTCCGTCCTACTCANGAGCTGCTCCATTTCTTCTGTCGAACC 2528
Db 1480 -----ProThrThrThrProSerProProThrThrThrThr-----P 1494
QY 2529 CTGCTCTATCAGCTGAACCTCTTTCTGGAGTGTAGTAGTACCCGCTCTCCCGAGC 2588
Db 1494 roProProThrThr-----:ProProSerP 1502
QY 2589 CCCTCAGCTGGTGGGCTGGGTGTGTCTCAGCGCAAAATGGGGCTCTGGTTCCAAATGGGCCA 2648
Db 1502 ro----- 1502
QY 2649 CTCTCATCTCTCTTGTTCCTTGTGCGAAGAACCTTTGTTCACTCCACTGCCCTCTCT 2708
Db 1503 -----PrometT 1505
QY 2709 AGTTCCCGACCTTTTCTCTCTGCTTCTCCCTGCCAAATTTCTCCAAGGAGTGTCTA 2768
Db 1505 hrThrProile-----T 1509
QY 2769 CACCTCTGCTCCACTTCTCTCT-----CCACCACACTCACTTCTTAACCCCTGCAATCTGG 2825
Db 1509 hrProProAlaSerThrThrLeuProProThr-----ThrThrPro----- 1523
QY 2826 CTTCCAGGCCCCAGCAATGGTTCTCTCAAGGTCCTCAGGACCTCTTGGCCAGCCGGA 2885
Db 1524 -----SerProProThrThrThrThrThrThrThrProProProT 1536
QY 2886 CAGTGTGTTGAAGGCTCATCTCTCTGCTGTCTGTTTGGCAGCCACACTGCTGAGCGCTG 2945
Db 1536 hr----- 1536
QY 2946 CTGCCTTCTCGAATCTCTTCTTGGTCTCTGCACTCTCTGGGCGCACTTCTACCTCT 3005
Db 1537 -----ThrThrProSerProProThrThrThr 1545
QY 3006 CAGCTCTCTCAGGCT 3065
Db 1546 ProilerProProProThrSerThrThrThrLeuProProThrThrThrProSerPro--- 1564
QY 3066 GTTTCGCCACCCAGCCCAATCAGCACGCTCTCTCTGAGGCTGTGCGTCTCTCTCTCT 3125
Db 1565 -----ProProProThrThrThrThrProProProProProProProProPro 1582
QY 3126 CTTTTTCTACGCTCTCCA-----TTGGAGAGCTCACACCGCCACTGCTTCAAC 3176
Db 1583 ThrThrThr--ProSerProProThrThrThrThrThrThrThrProProThrThrThr 1602
QY 3177 TGTACCTGCATACAAATGATATCTTATTGAAAACTCAGGAGGCGCATGAAACAAGA 3236
Db 1602 roSerPro-----ProThrThrThrThrThrThrThrProProProThrThrThr 1617
QY 3237 AGCCTAGC---ATGGAGACAGGCGCAGTGTCTCAGGGGACACAAAAATAGAACTTTGGGA 3293
Db 1617 hrProSerProProThrThrThrThrThrThrThrProProThrSerThr--ThrThrLeu--- 1635
QY 3294 GCAGTATCTCTTGTGTGTGAGCCAGCGGCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCT 3353
Db 1636 -----ProProThrThrThrThrThrThrThrProProThrThrThrProSer 1642
QY 3354 CTTTTTCAGCTGAATCTAGTGGGCGCCCTTTTCACAGCGCAGACTTCAACCTCCGCC 3413
Db 1643 ProProProThrThrThrThrThrProPro-----ProThrThrThrThrPro 1657
QY 3414 GCCTGGGSCAGACTCTGCCAGGCGCACTGTTGACTTTGAGTGTGGAGTGTGGCGAAGCCTCA 3473
Db 1658 SerProProThrThrThrThrProSerProProileThrThrThrThrThrThrProPro 1677
QY 3474 ATGACAAACTGGGCTGACCCAGAACTACGAGGCGCTTACAGCCACCTTCTGTGTACTTGC 3533

Db 1678 ThrThrThrProSerSerProileThrThrThrProSer--ProProThrThrThrMetTh 1697
QY 3534 GTGGCTTCAACCGTCAGGCTGCCACTGCTGAGCTGGCGGAGCTGGCCCTTCTTCTGCA 3593
Db 1697 rThrProSerProThr-----ThrThrProSerSerProileTh 1710
QY 3594 CCAGCTCCAGGGCTGCTGGGCGACATTCGGGCGTTCATGGCAGCTCTGGGCTACCCAC 3653
Db 1710 rThrThrThrThrPro-----SerSerThrThrThr 1720
QY 3654 TGCCCCAGCCCTGCTGGGACT-----GAACCCACTTGGACTCTGGCCCTG 3701
Db 1720 r-ProSerProProProThrThrMetThrThrProSerProThrThrThrProSerPro- 1739
QY 3702 CCCAGTGAAGTCTCTCCAGAGATGAGCAGCTTCTGCTGCTGAAGAGCTGACAGCT 3761
Db 1740 ProThrThrThrMetTh-----ThrLeuProPro 1749
QY 3762 GGCTGTGGCGTCGCCCAAGGACTTCAACCG-CTCAAGAAAGATGAGCCTCCAGCA 3820
Db 1750 ThrThrThrSerSerProLeuThrThrThrProLeuProProSerileThrProPro--- 1768
QY 3821 GCTGAGTCACTGACCTGGGCTCATGGCTCTGACTTCTGACTTCTCTCTCTCTG 3880
Db 1769 -----ThrPheSer----- 1771
QY 3881 CTCCTCCCTTCAACCTCTGCTCCACTTTGTGAGAGCAGCCCTGTATGTCACACACTGT 3940
Db 1772 ---ProPheSerThrThrThrPro-----ThrThrProCysValPro----- 1784
QY 3941 GAGCCAGGAGACAGAAAGCTGTGAGCCTCTGGCCCTTCTCTGACCGGCTGGCGGTGTGAT 4000
Db 1785 -----LeuCysAsnTrpThrGlyTrp--LeuAsp 1793
QY 4001 GCGATCAGCCCTGTCTCTCCCTCCCTCCCTCCAAAGTCTA----- 4039
Db 1794 SerGlyLysProAsnPheHisLysProGlyGlyAspThrGluLeuileGlyAspValCys 1813
QY 4040 ---CCGAGCTCGGGAGGA----- 4054
Db 1814 GlyProGlyTPAlaAlaAsnIleSerCysArgAlaThrMetTyrProAspValProile 1833
QY 4055 GGTACAGTAGGCGCTCTGCTCTCTCTTCTACAGAAAGTCACTGTCGAGGAGGTGAA 4114
Db 1834 GlyGlnLeuGlyGlnThrValValCysAspValSerValGlyLeuileCysLysAsnGlu 1853
QY 4115 GTGGTTCAGGTTGGTGCAGAGCGCTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4156
Db 1854 AspGlnLysProGlyValileProMetAlaPheCysLeuAsnTyrGluileAsnVal 1873
QY 4157 -----TTGCCTACCTTGGCCAGTG-----CCC 4180
Db 1874 GlnCysCysGluCysValThrGlnProThrThrMetThrThrThrThrGluAsnPro 1893
QY 4181 ACCCAGCCCTCAGGTGGCAGCATCTGGAGGCGAGGGGTGAGGGGCCACCCACACATG 4240
Db 1894 ThrProProThrThrThrProileThrThrThrThrValThrProPro----- 1911
QY 4241 CCTTTCTGGGTGAAGCCCTTTGGCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4300
Db 1912 -----ThrProThrGlyThrGlnThrProThr 1920
QY 4301 TCCCCCAATCACTATATATATCCAAATTCAGGAAACAAACATGTCGCAATTTCTACAAA 4360
Db 1921 ThrThr-----ProileThrThrThrThrValThr----- 1931
QY 4361 AAGAGATGAGATTAAAGTGCAGGCTGGGTGCTGCTGCTGAGGTGCTTATAAACAGA 4420
Db 1931 ----- 1931
QY 4421 AGAGAAAATCTGAAAGACACAGGGCGAGGACAGACAGACAGCCAGGAGTCTCCAA 4480
Db 1931 ----- 1931

Qy	1856	TGCGGGGGGGGTTGGGTGGGACAGAGGGGCCCACTCCATCGCTGCCTTCAGCT	1914
Db	1332	-----AlaProGluAspIleGluCysArgSerVal	1341
Qy	1915	CGCCTCTGCCCCACAGACCTGGGGCCCTGCTCTCGAGCCAGGGCCCTCCCTTCGGTCT	1974
Db	1342	---Lys-----AspProHisLeuSerLeuGluGln	1350
Qy	1975	GCCTCTCCCATCCTAGCTGGGCTCTAGGGGGGTTCATGGGGAAGGAGC	2025
Db	1351	-----HisGlyGlnLysValGlnCysAspValSerVal	1361
Qy	2026	-----TGTAGGAAACCCAGGACAGTAGTCGACGGGGTTTAGGTGTGGATGAGAGT	2076
Db	1362	GlyPheIleCysLeuAsnGluAspGlnPheGlyAsnGlyProPheGlyLeuCysTyrAsp	1381
Qy	2077	TATGCTGTAAAGATN-----TGGGGGTGTCACAGAGGTCTTCAGAGACCCAGGAG	2127
Db	1382	TyrLysIleArgValAsnCysCysTrp--ProMetAspLysCysIleThrThrPro---	1399
Qy	2128	AGAAGAAAGAGGTTGGAGGACCGAGGCACCATGGGGAACCGGCCCTCTCTCCGCTG	2187
Db	1400	-----SerProProThrThrThrProSerProProProThrThrThrThr	1414
Qy	2188	TTCTCTTCCACATCCAGACCTACTCTGGAGCCAGGAAAGAAAGGGAAGAGGTGG	2247
Db	1414	hrThrLeuPro-----	1417
Qy	2248	CGGGGAGCTGGCTCCAGCCCCAGGATACACCGAGGAAATAGTTGTCTCTGTGCTTGT	2307
Db	1418	--ProThrThrThrProSerProProThrThr-----	1428
Qy	2308	CAGCGTGTGAACCTCCCCCTGGGCCCTTGCCATATCCAGG-	2348
Db	1428	hrThrThrThrProProProThrThrThrProSerProProIleThrThrThrThrThr	1448
Qy	2349	CTCTCCCTTGCTCTCCCTTTTCCAGTTATACATCTCCCTCATCCCTTTCCCTGGG	2408
Db	1448	roLeuProThrThrThrProSerProProIleSerThrThrThrProProProThrThr	1468
Qy	2409	CCCCAGCGCTCCCGAGGGTTGGAAAGGGCTCTGCCCTCTCCCTATACCATGCTGTC	2468
Db	1468	hrThrProSerProPro-----ThrThrThrProSerPro-----	1479
Qy	2469	TTCCATAGCTTCCTCCTCCTCACTCATGAGACTGCCTCCATTCTTCTCTCTGCAACC	2528
Db	1480	-----ProThrThrThrProSerProProThrThrThrThrThrThr-----	1494
Qy	2529	CTGCTCTATCAGCTGAACCCCTCTCTTCGAGGTGTGTAGTAGTACCCGTCTCTCCAGC	2588
Db	1494	roProProThrThr-----ThrProSerP	1502
Qy	2589	CCCTCAGCTGGTGGGCGCTGGGTGTGTACAGGGCAATGGGGCTCTGGTCCCATGGGCCA	2648
Db	1502	ro-----	1502
Qy	2649	CTCTCATCTCTCTTGTTCCTTGTCAGAAAAACCTTTGTCTTCACTCCAGTGCCTCTCT	2708
Db	1503	-----ProMetT	1505
Qy	2709	AGTTCGACCCCTTTTCTCTCTGCTTTTCCCTGCCAAATTTCTCCAAGGAGTGCTTA	2768
Db	1505	hrThrProIle-----	1509
Qy	2769	CACCTCTGCTCTCCACTTCTCTCT--CCAGCCACTCACTTCTTAAAGCCCTGCAATCTGG	2825
Db	1509	hrProProAlaSerThrThrThrLeuProProThr-----ThrThrPro-----	1523
Qy	2826	CTTCCAGGCCCCAGCAATGGTTCTCTCCAAGGTGTCAGGACCTCTCTTGCACAGCCCGA	2885
Db	1524	-----SerProProThrThrThrThrThrThrThrThrProProProPro	1536
Qy	2886	CAGTGTTTTGAAGGTCATTCCTCTGCTGTCTGTTTTGCAGCCACCACTGCTCAGGGCTG	2945

Db	1536		hr-----	1536
Qy	2946	CTGCCCTCTCGA	ACTCCCTCTCTGGTCTCTGCACTCTCTGGCCACCTTCTACCTCT	3005
Db	1537	-----	-----Thr-ThrProSerProProThrThrThr	1545
Qy	3006	CCAGCTCTCTCCAG	GGCTCCTCTCTCTGCTGCTGCCCCACAGCGGGCACTCTCCCAAG	3065
Db	1546	ProIleThrProPro	ThrThrLeuProProThrThrThrProSerProPro	1564
Qy	3066	GTTTGGCCACCGAC	CAATACGACCTCCTTCTGAGCGCTTGTTGGGTCTCTCTCTCTCT	3125
Db	1565	-----	-----ProProThrThrThrThrProProThrThrThrProSerProProThr	1582
Qy	3126	CCTTTTCTACGCC	CTCCA-----TTGGAGAGCTCACACCGCCACTGCTTCAAC	3176
Db	1583	ThrThrThr- -	ProSerProProThrIleThrThrThrThrProProProThrThrThr	1602
Qy	3177	TGTCACCTGCATAC	AAATGATATCCTTATTGGAATAACTCAGGAGGCCATGAACAAAGA	3236
Db	1602	roSerPro-----	-----ProThrThrThrThrThrProProProThrThrThr	1617
Qy	3237	AGCCTAGC- -	-ATGGAGACAGGGCCAGTGTACGGGACACAAAATAAGAACTTTGGGA	3293
Db	1617	hrProSerProPro	ThrThrThrProIleThrProProThrSerThr- ThrThrLeu- -	1635
Qy	3294	GCAGGTATCTCT	TGTTGGTGTAGCCAGCGCTCGCCCTCTCTCTCTCTCTCTCTCTCTCT	3353
Db	1636	-----	-----ProProThrThrThrThrProSerPro	1642
Qy	3354	CCTTTTACAGCTGA	ACTACCTGGGCCCCCTTTCAACGAGCCAGACTTCAACCCCTCCCC	3413
Db	1643	ProProProThr	ThrThrThrThrProPro-----ProThrThrThrThrPro	1657
Qy	3414	GCTGGGGCAGAGAC	TCTGCCCAGGGCCACTGTTTCATTTGGAGTGTGGCGAAGCCCTCA	3473
Db	1658	SerProProThr	ThrThrThrProSerProProIleThrThrThrThrThrProProPro	1677
Qy	3474	ATGCAAACTGCGGT	AGCCAGACTAGAGCCCTACAGCCACCTTCTGTGTACTTGC	3533
Db	1678	ThrThrThrProSer	SerProIleThrThrThrProSer- -ProProThrThrThrMetTh	1697
Qy	3534	GTGGCTCAACCGT	CAGGCTGCCACTGCTGAGCTGCGCGCAGCCTGGCCACTTCTGCA	3593
Db	1697	rThrProSerPro	Thr-----ThrThrProSerSerProIleTh	1710
Qy	3594	CCAGGCTCCAGGCG	CTGTGGGCGAGCATTCGGGGCGTCATGGCAGCTCTGGGCTACCCAC	3653
Db	1710	rThrThrThrPro	-----SerSerThrThrThr	1720
Qy	3654	TGCCCCAGCGCTG	CTCTGGGACT-----GAAACCCACTGGACTCTCTGGCCCTG	3701
Db	1720	r-ProSerProPro	ThrThrMetThrThrProSerProThrThrThrProSerPro-	1739
Qy	3702	CCACAGTCTACTT	CTCTCCAGAAGATGGACGACTTCTGGCTGCTGAAGAGCTGCAGACCT	3761
Db	1740	ProThrThrThr	MetThr-----ThrLeuProPro	1749
Qy	3762	GCGTGTGGCGCT	CGGCCAAGGACTTCAACCGG-CTCAAGAAGAAGATGCAGCCTCCAGCA	3820
Db	1750	ThrThrThrSer	SerProLeuThrThrProLeuProProSerIleThrProPro-	1768
Qy	3821	GCTGCAGTCA	CCCTGCACCTGGGGGCTCATGGCTTCTGACTTCTGACCTTCTCTCTCTG	3880
Db	1769	-----	-----ThrPheSer-----	1771
Qy	3881	CTCCCCCTTCAAA	CCCTGCTCCCACTTTGTGAGAGCCAGCCCTGTATGCCAACACACTGTT	3940
Db	1772	---ProPheSer	ThrThrPro-----ThrThrProCysValPro-----	1784
Qy	3941	GAGCCAGGACAGA	AGCTGTGAGCCCTCTGGCCCTTCTCTGAGCCGCTGGGCGTGTGAT	4000

[illegible]

Db	1468	hrThrProSerProPro	:::			ThrThrThrProSerPro	1479
Qy	2469	TTCCATAGCTTCCTCCCTGCTCTACTCATGAGACTGCCTCCATTTCTCTCTCTGCAACC					2528
Db	1480	ProThrThrThrProSerProProThrThrThrThrThr				ThrThrThrThr	1494
Qy	2529	CTGCTCCATCAGCTGAACCTTCTTTCCGAGGTGTAGTGAGTACCCGCTCTCTCCCGCAGC					2588
Db	1494	roProProThrThr				ThrProSerP	1502
Qy	2589	CCCTCAGCTGTGGCGCTGGGTGTGCACGGCAATGGGGCTCTGGTTCCAATGGGCCA					2648
Db	1502	ro					1502
Qy	2649	CTCTCATCTCTCTTGTTCTCTGTGTGCAGAAAAACCTTTGCTTCACTCCAGTGCCTCTCT					2708
Db	1503					Promet	1505
Qy	2709	AGTTCCCGACCTTTTCTCTCTCTGCTTTCCCTGCAGAAATTTCTCAGAGAGTGGTCTA					2768
Db	1505	hrThrProIle					1509
Qy	2769	CACCTCTGCTCCACTTCTCTCT---CCACCACTCACCTTCTTAACCCCTGCAATCTGG					2825
Db	1509	hrProProAlaSerThrThrLeuProProThr				ThrThrPro	1523
Qy	2826	CTTCAGGCCCCAGCAATGGTTCTCTCCAAGTCTGTCAGGCACCTCTTGTCCCAAGCCGA					2885
Db	1524					SerProProThrThrThrThrThrProProPro	1536
Qy	2886	CAGTGTTTGAAGGCTCATTCCTCTGCTGTGTTTGCAGCCACACTGCTGAGCGCTG					2945
Db	1536	hr					1536
Qy	2946	CTGCCCTCTCGAACTCTCTTCTCTGGTCTCTGCATCTCTCTGGGCGACCTTCTACCTCT					3005
Db	1537					ThrThrProSerProThrThrThr	1545
Qy	3006	COAGCTCTCCAGGCTCCTTCTCTCTCTGCTGCTGCCCCACAGCGGGCACTCTCCCAAG					3065
Db	1546	ProIleThrProProThrSerThrThrThrLeuProProThrThrThrThrProSerPro				ThrThrThrThrThrThr	1564
Qy	3066	GTTTGCCCAACCAGCCCAATCAGACGTCCTCTCTGAGCGTCTGTGCGTCTCTCTCTCTCT					3125
Db	1565	ProProThrThrThrThrThrProProProThrThrThrThrProSerProProThr					1582
Qy	3126	CCTTTTCTACGCTCTCCA-----TTGGAGAGCTCAACCACCGCCACTGCTTCAAC					3176
Db	1583	ThrThrThr--ProSerProProThrIleThrThrThrThrProProProThrThrThr					1602
Qy	3177	TGTCACTGTCATACAATGATATCTTATTTGAAAAAATCTCAGGAGGCCATGAACAAGA					3236
Db	1602	roSerPro-----ProThrThrThrThrThrProProProProThrThrThr					1617
Qy	3237	AGCCTAGC--ATGAGACAGAGCCAGTCTCAGGGGACACAAAAATAGAACTTTGGGA					3293
Db	1617	hrProSerProProThrThrThrProIleThrProProThrSerThrThrThrLeu					1635
Qy	3294	GCAGGTATCTCTCTGGTGGTGAGCCAGCGGCTCTGCCCTCTCTCTCTCTCTCTCTCTCT					3353
Db	1636					ProProThrThrThrProSer	1642
Qy	3354	CCTTTTCAAGCTGAATCACTGGGGCCCTTTCAACAGCCAGACTTCAACCTCTCCC					3413
Db	1643	ProProProThrThrThrThrProPro-----ProThrThrThrThrThr					1657
Qy	3414	GCCTGGGGCAGAGACTCTGCCCGAGGCCACTGTTGACTTGGAGGTGTCGCGAAGCTCA					3473
Db	1658	SerProProThrThrThrThrThrProSerProProIleThrThrThrThrThrProPro					1677
Qy	3474	ATGACAAATCGCGGTGACCCAGAACTACGAGCGCCTACAGCCACCTTCTGTATTATTC					3533


```

; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
; US-09-919-497-56

```

Alignment Scores:	
Pred. No.:	4.77e-06
Score:	233.00
Length:	1806
Percent Similarity:	28.00%
Conservative:	90
Best Local Similarity:	21.61%
Mismatches:	477
Query Match:	2.47%
Indels:	555
DB:	10
Gaps:	73

US-09-931-704-3 (1-5087) x US-09-919-497-56 (1-1806)

944	QY	CCCTCAGCTGCAGTGGGACTGGAAGACTGGCAGGAAGCTAGGTGTAACA	CTGGACACAG	1003
449	Db	ProAlaGlyIleMetGlyProProGlyLeuGlnGlyProThrGlyProProGlyAspPro	468	
1004	QY	GCAGGTGCAGCTGCAGTCCCTAGGCTGGC---	-----CCGCTCCCT	1042
469	Db	GlyAspArgGlyProProGlyArgProGlyLeuProGlyAlaAspGlyLeuProGlyPro	488	
1043	QY	CCATGTACACACATATCATGTTGGCCACACACACAGTGGCCACATCCCAAAGACTCTCT	1102	
489	Db	ProGlyThrMetLeuMetLeuPro-PheArgTyrGlyGlyAspGlySerIysGlyPro--	507	
1103	QY	CAGCTGCACACACAGATCCATTCTCAAGTATCTACTGATAGACATCATCGTGCCCAATC	1162	
508	Db	-----ThrIleSerAlaGlnGluAl	514	
1163	QY	CTCATCTCAAACATACACATACATCCCTCTCTTTCTCTCCCGTCTTCCAGGAGTGTTC	1222	
514	Db	aGlnAlaGlnAlaIle-----	-----LeuGlnGlnAlaArgIleAlaLeuArgG	529
1223	QY	TGCTTCATCCCTCTCTGCTCCATCTGGTGTCCCAACCTCATCCACCCCAACAGGCCCAAGG	1282	
529	Db	yProProGlyProMet-----	-----GlyLeuThrGlyArgProGlyProVa	543
1283	QY	TGGGGACAGACACCTGAGGGGTGCAGCTCTCCCGTGTGGCCCGCGCGCTCA	1342	
543	Db	lGly-----	-----GlyProGlySerSerG	550
1343	QY	TGCTTCTGTCATCTGTCACACAGGGACTCGTGGGGATGTTAGCGTGCCTGTGCACG	1402	
550	Db	yAlaIysGlyGluSerGlyAspProGlyProGlnGly-----	-----	562
1403	QY	GTGCTTGGCACCTCCCTGCAGTGCACGCTCTCAATCGCACAGGGGACCCAGGCGCTGGC	1462	
563	Db	-----	-----ProArgGlyValGlnGlyProPr	570
1463	QY	CCCTCCATCCAGAAAACCTTATGACCTTCACCCGCTACTTGGAGCACCACTCCCGCAGCTTG	1522	
570	Db	oGlyProThrGlyLysProGly-LysArgGlyArgProGlyAlaAsp-----	-----G	586
1523	QY	GCTGGACCTATGTGTAGTATCCAGCGTAGGAATCTGGGAGTGTGGGAGGAGTGHAGGATT	1582	
586	Db	ly-GlyArgGlyMetPro-----	-----GlyGluProGlyAlaLys	597

Db	871	YValAlaGlyLysProGlyProArgGlyGlnArgGlyProThrGlyProArgGlySerAr	891
QY	2597	TGCTGGCCCTGGTGTGTCAGCGGCAAAATGGGGC-----TCTGGTTCCAAATGG	2644
Db	891	gGlyAlaArgGly---ProThrGlyLysProGlyProLysGlyThrSerGlyGlyAspG	910
QY	2645	GCCACTCTCATCTCTCTTGTCTTGTGTGCAGAAAACCTTTGCTTCACTCCACTGCCT	2704
Db	910	yPro-----	911
QY	2705	CTCTAGTTCCGACCCCTTTTCTCTCTGGCTTTCCTGCGCAAAATTTCTCCAGGAGTGG	2764
Db	912	-----ProGlyProProGlyGluArgGlyProGlnGlyPr	923
QY	2765	TCTACACCTCTGCCT---CCACTCTCTCCACCACCTCACCTTCTTAACCCCTGCAAT	2821
Db	923	oGlnGlyProAlaGlyPheProGlyProLysGlyPro-----	935
QY	2822	CTGGCTTCCAGGCCCCAGCAATGTTCTCTCCAAAGTCTGTCAGGCACCTCTTCCCAAGC	2881
Db	936	-----ProGlyPro-----ProGlyArgMetGlyCysProGlyHisPr	948
QY	2882	CCGACAGTGTTTGAAGGCTCAATTCCTTCTGCTGTCTGTTTTCGAGCCACACTGCTGAGC	2941
Db	948	oGlyGlnArgGlyGluThrGlyPheGln-----	957
QY	2942	GCTGTCTTCTCGAACT---CCTCTCTTGGTCTCTGCACCTCTCTCT---GGGCC	2992
Db	958	-----GlyLysThrGlyProGlyProGlyGlyValGlyProGlnGlyPr	974
QY	2993	ACCTTCTACTCT---CCAGCTCTCTCCAGGCTCTCTTC-----	3028
Db	974	oThrGlyGluThrGlyProIleGlyGluArgGlyTyrProGlyProGly-----	991
QY	3029	CTCTGTCTCTGCCCCACAGCGGCACCTCTCCAAAGTTTGCACCCAGCCAATCAGC	3088
Db	992	-----ProGlyGluGlnGln-----	997
QY	3089	ACGTCTCTCTCTGAGCG---TCTTGTGGTCTCTCTCTCTCTCTCTTTTCTACGCC	3139
Db	997	yLeuProGlyAlaAlaGlyLysGluGlyAlaLysGlyAspProGlyProGlnGly-----	1015
QY	3140	TCTCCATTGGAGAGCTCACCGCCACTGCTTCAACTGTCTCAGCTGCATACAAATGATAT	3199
Db	1015	-----	1015
QY	3200	CCTTATTTGAAAA---ACTCAGGAGGCCATGAACAAGAAGCCTTACG	3244
Db	1016	-IleSerGlyLysAspGlyProAlaGlyLeuArgGlyPheProGlyGluArgGlyLeuPr	1035
QY	3245	ATGGAGACAGGCCAG-----TCTCAGGGACACAA-----	3275
Db	1035	oGlyAlaGlnGlyAlaProGlyLeuLysGlyGlyGlyProGlnGlyProProGlyPr	1055
QY	3276	-----AAAAATAGAAACTTTGGGAGCAGGTATCTCTTGGTGTGAGCCAGCGCT	3325
Db	1055	oValGlySerProGlyGluArgGlySerAlaGlyThrAlaGlyProIleGlyLeuArgG	1075
QY	3326	CTGCCCTCTCTTCCCCATCACCTCTCTCTTTTCACAGCTGAACCTACCTGGGCCCT	3385
Db	1075	yArgProGlyProGlnGlyProProGlyProAlaGlyGlu-Lys-----GlyAlaProG	1093
QY	3386	TTCAACGAGCCAGACTTCAACCTCTCCGCTCGGGGACAGACTCTGCCAGGCCACT	3445
Db	1093	lyGluLysGly-----ProGlnGlyProAlaGlyArgAspGlyValGlnGlyProV	1110
QY	3446	GTTGACTTGGAGTGTGGCGAAGCCTCAATGACAAACTCGCGCTGAC-----	3492
Db	1110	alGlyLeuProGlyProAlaGlyProAlaGlySerProGlyGluAspGlyAspLysGlyG	1130
QY	3493	-----CCAG	3496

QY 4400 G-----GAGGTGCCCTATAAACCAAGAGAGAAATACTGAAGCACAGGGGCGGAGCA 4453
Db 1425 aGlyGlnAspGlyProGlyPro-----MetGlyProProGlyLeuProG1 1441
QY 4454 GACCAGACAGACCCAGGAGTCTCCAAAGCACAGAGTGGCAACAAACCCAGGCTGAGC 4513
Db 1441 yLeuLysGlyAspProGly---SerLysGlyGlyLysGlyHisProGlyLeuLysGlyLe 1460
QY 4514 ATCAGGACCTTCCGCGAATTGCTTCCAGTATTACGGTGCCTCTTCTCTGCCCCCTTTC 4573
Db 1460 uileGlyProProGlyGluGlnGly-----G1 1469
QY 4574 CCAGGTATCTGTGGTTGCCAGGCTGGGAGGCGCAACCATAGCCACACAGGATTTTC 4633
Db 1469 uLysGlyAspArgGlyLeuProGlyThrGlnGlySerPro-----1482
QY 4634 CTGAAGTTTACAAATGCAGTAGCATTTTGGGTGTAGGT---GGCAGCTCCCAAGGCC 4690
Db 1483 -----GlyAlaLysGlyAspGlyGlyLeuProGlyPr 1493
QY 4691 CTGC-----CCCGCAGCCGCCACC 4709
Db 1493 oAlaGlyProLeuGlyProProGlyProPro 1503

RESULT 14

US-09-735-367B-3
; Sequence 3, Application US/09735367B
; Patent No. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Cairra, Francoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2005
; TYPE: PRT
; ORGANISM: mammal
US-09-735-367B-3

Alignment Scores:
Pred. No.: 6,6e-06 Length: 2005
Score: 230.50 Matches: 387
Percent Similarity: 32.03% Conservative: 212
Best Local Similarity: 20.70% Mismatches: 619
Query Match: 2.44% Indels: 654
DB: 10 Gaps: 90

US-09-931-704-3 (1-5087) x US-09-735-367B-3 (1-2005)

QY 15 CTGGCGGATGGGATATTAAAGCTTCGCGGAGCGCGCTCGCCCTCCCACTCCGCCA 74
Db 257 ProProGlyTrpAsn-----GlnLeuProSerGlyAlaLeuGlnPro-----ProPro 272
QY 75 GCGTCCGGG-----AGAGGAGCGCCGCCCGGC-----CGGCCCGAGCCCC 113
Db 273 AlaGlnGlySerLeuGlyThrMetThrAlaAsnGlnGlyTrpLysAlaProLeuPro 292
QY 114 AGCCCATGGACCTCGAGCAGGTGAAACCCAACTAGCCCTGCTCTTCATAACATGA 173
Db 293 GlyProMetGlnGlnGlnGlnAlaArgProSerLeuAlaThrVal-----308
QY 174 CAAGCAGCGCCCATCTGATACCTAAACCGACCAAGTCACAGCCCTCCCACTCACCTCT 233
Db 309 -----Gln-ThrProSerHisProProProProProPro 320

QY 234 GCTGTGCCAG-----243
Db 320 eGlySerGlnGlnAlaSerGlnAlaHisThrAsnPheProGlnMetSerAsnProGlyG1 340
QY 244 ----ACCTCACCATCTTGTGGACTCAAA-----270
Db 340 nPheThrAlaProGlnMetLysSerLeuGlnGlnGlyGlyProSerArgValProThrProLe 360
QY 271 -CCTCAACCCGCTATAATCAACCAATCCCAAGTCTAAACTTAATCTGAACCTTTTAAAGT 329
Db 360 uGlnGlnProHisLeuThrAsnLysSerProAla-----SerSerProSerSerPh 377
QY 330 AACCCAGTCTTAAACCTAACCTAGCCCAATGCCAATATATATCTACCTAGCCCAACCTC 389
Db 377 eGlnGlnGlySerProAlaSerSerProThrValAsn-----GlnThrG1 392
QY 390 AACTGCCCTTTGCCAGTCCAAAGTCTCCACTGAATCCTCACCTTGGTCTCACTGAAATC 449
Db 392 nGlnGlnMet---GlyProArgProProGlnAsnProLeu---ProGln---GlyPh 409
QY 450 CCAGAAAGCATATTTCCCACTGCCCACTCCCTCTTACAGCACCCCAACCTCGCCCTC 509
Db 409 eGlnGlnProValSerSerProGlyArgAsnProMetValGlnGlnGlyAsnValProPr 429
QY 510 TGGACTCTCGTATCTCTGGGATGTCAAACTCTGAGTGCCATCAGCCAAAGCCCGAC 569
Db 429 oAsnPheMetVal-----MetGlnGlnProProAsnGlnGlyPr 443
QY 570 TCGTCAATGCACCTCTCT-----CCCTTCTGTCCCACTTTCAGGCTGA 617
Db 443 oGlnSerLeuHisProGlyLeuGlyGlyMetProLysArgLeuProProGlyPheSerAl 463
QY 618 TGAAGAGCGCTCATGAAGTCCAACTTTTCCCACTAACCAAGAACGCGG---GTGAA 674
Db 463 aGlyGlnAlaAsn-----ProAsnPheMetGlnGlyGlnValPr 476
QY 675 CTCCACACTGCCACCGTTCCTGAGAGTGAGCACTAAATCTCTTCAATCTAACCCAC 734
Db 476 oSerThrThrAlaThrThrProGlyAsnSerGlyAlaProGlnLeuGlnAlaAsnGlnAs 496
QY 735 CTACACTTCCCACTCAGGAATCACAATCTAGATATATACCAAACTAAG-----786
Db 496 nValGln-----HisAlaGlyGlyGlnGlyAlaGlyProProGlnAsnGlnMetGlnVa 514
QY 787 -CCCATAGGCGACCGCCCTAGTGTCTAACCCCTATACCTTCTCTTCTTCTGAGT 845
Db 514 lSerHis---GlyProProAsnMet-----MetGlnPro 524
QY 846 TCTGTTCTTGGC-----GGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 899
Db 525 SerLeuMetGlylLeHisGlyAsnMetAsnGlnGlnAlaGlyThrSerGlyValPro 544
QY 900 CAGCTGCGAGCTCTGACATGTGTCTCCACCTCTGACTCCCTCCCTCAAGCTGACGTG 959
Db 545 GlnValAsnLeuSerAsnMetGlnGlyGlnProGlnGlnGlyProProSerGlnLeuMet 564
QY 960 GGACTG---GAAGACTGCGAAGCTAGGTACAGTGGAAACACAGGAGGCTGACCTG 1016
Db 565 GlyMetHisGlnGlnLeValProSerGlnGlnMetValGlnGlnGlnGlnGlnGlnGln 584
QY 1017 CAGTCCC-----TAGCCCTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1055
Db 584 uAsnProGlnAsnProMetlLeLeuSerArgAlaGlnLeuMetProGlnGlnMetMe 604
QY 1056 TATACATGTTGGGACACACACAGTGGGACACATGCAAGAGCTCTCTCAGCTGACACACA 1115
Db 604 tValAsnProProSerGlnAsnLeuGlyProSerProGlnArgMetThrProProlySgl 624
QY 1116 GATCCATCTCAAGTATCTACTGATAGACACTCATGCGTGGCCAGCTCTCAT-----1167
Db 624 nMetLeuSerGlnGlnGlyProGln-----MetMetAlaProHisAsnGlnMe 640
QY 1168 -----CCTCAAAACATACATGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1220


```
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2063
; TYPE: PRT
; ORGANISM: Human
; US-09-735-367B-2

Alignment Scores:
Pred. No.: 1,18e-05 Length: 2063
Score: 226.00 Matches: 387
Percent Similarity: 32.03% Conservative: 212
Best Local Similarity: 20.70% Mismatches: 619
Query Match: 2.40% Indels: 654
DB: 10 Gaps: 90

US-09-931-704-3 (1-5087) x US-09-735-367B-2 (1-2063)
Qy 15 CTGCGGATGGATTATTAAGCTTCGCGGAGCGCGGCTCGCCCTCCCACTCCGCCA 74
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 315 ProProGlyTrpAsn-----GlnLeuProSerGlyAlaLeuGlnPro-----ProPro 330
Qy 75 GCCTCCGG-----AGAGGAGCGCACCCGGC-----CGGCCGAGCCCC 113
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 331 AlaGlnGlySerLeuGlyThrMetThrAlaAsnGlnGlyTrpLysLysAlaProLeuPro 350
Qy 114 AGCCCATCGAGCTCCGAGCGAGGTGAACAAACCAACCTAGCCCTGCTCTTCAATACATGA 173
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 351 GlyProMetGlnGlnLeuGlnAlaArgProSerLeuAlaThrVal-----366
Qy 174 CAAGCAGCGCCCATCTGATACCTAAACGACCAAGTCACAGCCCTCCCACTCACCCCTCT 233
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 367 -----Gln-ThrProSerHisProProProProProPh 378
Qy 234 GCCTGCCAG-----243
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 378 eGlySerGlnGlnAlaSerGlnAlaHisThrAsnPheProGlnMetSerAsnProGlyGl 398
Qy 244 -----ACCTCACCACATCTTGTGGACTCAA-----270
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 398 nPheThrAlaProGlnMetLysSerLeuGlnGlyGlyProSerArgValProThrProLe 418
Qy 271 -CCTCAACCGCACTAAATCAACCAATCCCAAGTCTAAACTTAATCTGAACCTTTTAAAGT 329
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 418 uGlnGlnProHisLeuThrAsnLysSerProAla-----SerSerProSerSerPh 435
Qy 330 AACCAGTCTTAAACCTAACCTAGCCCAATGCCAATATATATACCTAGCCCAACCT 389
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 435 eGlnGlnGlySerProAlaSerSerProThrValAsn-----GlnThrGl 450
Qy 390 AACTGCTTTGGCAGTCCAAAGTGTCCATCTCACTTGGTCTCCTCACTCACTGAAATC 449
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 450 nGlnGlnMet---GlyProArgProProGlnAsnAsnProLeu---ProGln---GlyPh 467
Qy 450 CCAGAAAGCATATTTCCCACTGCCACATCCCTCTTACAGACCCCAACCTGGCCCTC 509
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 467 eGlnGlnProValSerSerProGlyArgAsnProMetValGlnGlnGlyAsnValProPr 487
Qy 510 TGGACTCTCTGGTATCTGGGATGTCACAACTCTGCGAGTGCATCAGCCCAACAGCCCGAC 569
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 487 oAsnPheMetVal-----MetGlnGlnProProAsnGlnGlyPr 501
Qy 570 TCGTCAATATGCACCTCTCT-----CCCTTCTCTGCCACCTTGCAGGCTGA 617
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 501 oGlnSerLeuHisProGlyLeuGlyGlyMetProLysArgLeuProProGlyPheSerAl 521
Qy 618 TGAAGAGGCTCATTTGAAGTCCAACTTTTCCCACTTCCCACTCAACCAAGAACCGG---GTGAA 674
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
```


Db	1465	SerAspProAsnLysLeuProSerValGlu-----						-----GluAsn	1476
Qy	3281	AGAAACTTTGGGAGCAGGATATCTCCTTGGTGTGACCCAGCGCTCGCCCTCCTCCTTC						-----GluAsn	1476
Db	1477	LysAsnLeuValSer-----						-----ProAlaMetArg	1485
Qy	3341	CCCATCACCCCTCTCCTTTTCACAGCTG-----						-----AACTAC	3373
Db	1486	GluAlaProThrSerLeuSerGlnLeuLeuAspAsnSerGlyAlaProAsnValThrIle						-----	1505
Qy	3374	CTGGGCCCCCTTTCAACAGCAGACTTCAACCTCCCGCTGGGGCAGAGACTCTG						-----	3433
Db	1506	LysProGlyLeuThrAspLeuGluValThrProValValSerGlyGluAspLeu						-----	1525
Qy	3434	CCACGGCCACTGT-----						-----	3478
Db	1526	LysLysAlaSerValIleProThrLeuGlnAspLeuSerSerLysGluProSerAsn						-----	1545
Qy	3479	AAACTCGCGCTACCCAGAACTACGAGGCTACAGCCACTCTCTGTGTGA-----						-----	3534
Db	1546	SerLeuAsnLeu-----						-----	1560
Qy	3535	TGGCTCAACCGCTCAGGCTGCGACTGCTGAGCTGCGCGCAGCGCTGCCACTTCGCA						-----	3594
Db	1561	HisProGluSerGluValSerAsnValAlaProSerIleProValMetSer						-----	1580
Qy	3595	CAGCCT-----						-----	3618
Db	1581	ArgProValSerSerSerIleSerThrProLeuProProAsnGlnIleThrValPhe						-----	1600
Qy	3619	-----						-----	3660
Db	1601	ValThrSerAsnProIleThrThrSerAlaAsnThrSerAlaAlaLeuPro-----						-----	1617
Qy	3661	GCCGTGCTGGGACTGAACCACTTGACTCC-----						-----	3714
Db	1618	-----						-----	1631
Qy	3715	CCTCACAAGATGGACACTTCTGGCTGCTGAGGAGCT-----						-----	3753
Db	1632	ProAsnAlaGlySerLysValMetValSerGluGlnSerAlaAlaGlnSerAsnAla						-----	1651
Qy	3754	-----						-----	3780
Db	1652	ArgProGlnPheIleThrProValPheIleAsnSerSerSerIleIleGlnValMetLys						-----	1671
Qy	3781	GGACTTCAACCGCT-----						-----	3813
Db	1672	GlySerGlnProSerThrIleProAlaAlaProLeuThrThrAsn-SerGlyLeuMetPr						-----	1691
Qy	3814	TCCACGACTGCAGTACC-----						-----	3850
Db	1691	oProSerValAlaValGlyProLeuHisIleProGlnAsnIleLysPheSerSerAl						-----	1711
Qy	3851	-----						-----	3858
Db	1711	aProValProProAsnAlaLeuSerSerSerProAlaProAsnIleGlnThrGlyArgPr						-----	1731
Qy	3859	ACTTCTGACCTTCTCTCTTCTGCTCCCTTCAACCCCTGCTCCCACTTTGTGAGAGCCA						-----	3918
Db	1731	oLeuValLeuSerSerArgAlaThrProValGlnLeuProSerProCysThrSerSe						-----	1751
Qy	3919	GCCCTGATGCCAACA-----						-----	3943
Db	1751	rProValValProSerHisProProValGlnGlnValLysGluLeuAsnProAspGlu						-----	1771
Qy	3944	-----						-----	3958
Db	1771	aSerProGlnValAsnThrSerAlaAspGlnAsnThrLeuProSerSerGlnSerThrTh						-----	1791
Qy	3959	-?-TGTGAGCCTCTGGCCCTTTCTTGAGCCGCTGGCGTGTGATCGCATCAGCCCTGT						-----	4014

1791	Db	rMetValSerProLeuLeuThrAsnSerProGlySerSerGlyAsnArgSerPro	Val	1811
4015	QY	CTCCCTCCCAACCTCCCAAGGCTACCGAGCTGGGAGGAGTACAGTAGGCGCTGTCT	CT	4074
1811	Db	lSerSer-----SerLysGlyLysGlyLysValAspLysIleG	L	1824
4075	QY	GTCTGTTTCTACAGGAAGTCATGCTCAGGGAGTGTGAAGTGTTCAGGTTCCGTGCAGA	A	4134
1824	Db	yGlnIleLeuLeuThrLysAlaCysLysLysValThrGlySerLeuGluLysGlyGlu	L	1844
4135	QY	G-----GCGCT	G	4140
1844	Db	uGlnTyrGlyAlaAspGlyGluThrGluGlyGlnGlyLeuAspThrThrAlaProGlyLe	L	1864
4141	QY	CATGGCC-----TCCTGCTCTTGTGCTACCACTTGCCAGTGCACACCGCC	C	4188
1864	Db	uMetGlyThrGluGlnLeuSerThrGluLeuAspSerLysThrProThrPro--ProAla	A	1883
4189	QY	CCTCAGGTGGCACA	C	4248
1884	Db	ProThrLeuLeuLysMetThrSerSerProValGlyProGlyThrAlaSerAlaGlyPro	P	1903
4249	QY	GGGTGAAGCCCTTTGGCTGCCCACTCTCCTTGATGGG--TGTTGTCTCCCTTATCCCC	C	4305
1904	Db	SerLeuProGlyGlyAlaLeuProThrSerValArgSerIleValThrThrLeuValPro	P	1923
4306	QY	AAATCACTCTATACATCC-----AATTCAGGAACAACATGGTGGCAATTCACACA	A	4359
1924	Db	SerGluLeuIleSerAlaValProThrThrLysSerAsnHisGlyIleAlaSerGlu	S	1943

Search completed: January 27, 2003, 17:22:54
Job time : 295.428 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:50:02 ; Search time 51.6061 Seconds
(without alignments)
5800.643 Million cell updates/sec

Title: US-09-931-704-3
Perfect score: 9432
Sequence: 1 aacctgcagtggtgctggc.....ccttgctgaagtcttctca 5087

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO spo01/US0931704/runat_27012003_154127_3671/app query.fasta_1.7189
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=ra1 -NINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0931704 @CGN 1.1.89 @runat_27012003_154127_3671 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	896	9.5	225	1	US-08-792-019B-2
2	896	9.5	225	1	Sequence 2, Appli
3	896	9.5	225	3	US-09-106-182-2
4	896	9.5	225	3	Sequence 2, Appli
5	863	9.1	225	1	US-09-016-534-2
6	863	9.1	225	1	Sequence 2, Appli
7	863	9.1	225	3	US-08-792-019B-5
8	863	9.1	225	3	Sequence 5, Appli
9	330.5	3.5	1078	3	US-09-988-819-5
10	330.5	3.5	1078	3	US-09-016-534-5
11	330.5	3.5	1078	4	US-08-963-825-21
12	330.5	3.5	1078	4	Sequence 21, Appl
13	330.5	3.5	1078	4	US-09-500-811-21
14	330.5	3.5	1078	4	Sequence 21, Appl
15	330.5	3.5	1078	4	US-09-570-573-21
16	330.5	3.5	1078	4	Sequence 21, Appl
17	330.5	3.5	1078	4	US-09-548-608-21
18	330.5	3.5	1078	4	Sequence 21, Appl
19	330.5	3.5	1078	3	US-08-931-820-4
20	330.5	3.5	1078	3	Sequence 4, Appli

13	318.5	3.4	2972	4	US-09-579-181-2	Sequence 2, Appli
14	318.5	3.4	3118	4	US-09-579-181-1	Sequence 1, Appli
15	305	3.3	1418	3	US-08-963-825-20	Sequence 20, Appl
16	305	3.3	1418	4	US-09-010-999-1	Sequence 1, Appli
17	305	3.3	1418	4	US-09-500-811-20	Sequence 20, Appl
18	305	3.3	1418	4	US-09-570-573-20	Sequence 20, Appl
19	305	3.3	1418	4	US-09-548-608-20	Sequence 20, Appl
20	302	3.2	1060	3	US-08-931-820-3	Sequence 3, Appli
21	298	3.2	870	2	US-09-010-928B-2	Sequence 2, Appli
22	296	3.1	1064	1	US-08-642-255-62	Sequence 62, Appl
23	295.5	3.1	1185	4	US-09-041-888-23	Sequence 23, Appl
24	295.5	3.2	1366	3	US-08-963-825-19	Sequence 19, Appl
25	295.5	3.2	1366	4	US-09-500-811-19	Sequence 19, Appl
26	295.5	3.2	1366	4	US-09-570-573-19	Sequence 19, Appl
27	295.5	3.2	1366	4	US-09-548-608-19	Sequence 19, Appl
28	295	3.2	745	2	US-09-010-928B-28	Sequence 28, Appl
29	295	3.2	1442	2	US-08-316-650-12	Sequence 12, Appl
30	295	3.2	1442	5	PCT-US95-02251-12	Sequence 12, Appl
31	293.5	3.1	1024	3	US-08-931-820-2	Sequence 2, Appli
32	293.5	3.1	1461	4	US-09-585-887-9	Sequence 9, Appli
33	293.5	3.1	1461	4	US-09-289-578-9	Sequence 9, Appli
34	291	3.1	1057	3	US-08-931-820-1	Sequence 1, Appli
35	289	3.1	822	4	US-09-219-849-49	Sequence 49, Appl
36	288	3.1	1958	1	US-07-945-283-2	Sequence 2, Appli
37	287	3.0	960	4	US-09-219-849-5	Sequence 5, Appli
38	286	3.1	1366	4	US-09-585-887-10	Sequence 10, Appl
39	286	3.1	1366	4	US-09-289-578-10	Sequence 10, Appl
40	279.5	3.0	897	1	US-08-397-633A-50	Sequence 50, Appl
41	278	3.0	1341	3	US-08-963-825-18	Sequence 18, Appl
42	278	3.0	1341	4	US-09-500-811-18	Sequence 18, Appl
43	278	3.0	1341	4	US-09-570-573-18	Sequence 18, Appl
44	278	3.0	1341	4	US-09-548-608-18	Sequence 18, Appl
45	277.5	3.0	1065	1	US-08-642-255-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-08-792-019B-2
; Sequence 2, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/792.019B
; APPLICATION NUMBER: US/08/792.019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-792-019B-2

Alignment Scores:
Pred. No.: 6,41e-57 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 1 Gaps: 2

US-09-931-704-3 (1-5087) x US-08-792-019B-2 (1-225)

QY 3311 GGTGAGCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACTGGGCCCCCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGACTCTGCCAGGCCCACTGTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGCGAAGCTCAATGACAACTGGCGTGAACCCAGAACTACGAGGCTACAGC 3514
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131
QY 3575 AGCTTGGCCCACTTCTGACACAGCTTCCAGAGGCTCTGGGCGAGCAATGCGGCGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTTGGGTACCCACTGCCCGAGCGCTGCCCTGGGAGTCAACCCACTTGGACTCCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
QY 3695 GCGCTTGGCCACAGTACTTCTCCAGAGAGATGGACGACTTCTGGCTGTGAAGGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTTGGCTGTGGCGCTCGGCCAAGAGCTTCAACCGCTCAAGAGAGATGCAAGCCT 3814
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCAGTACCTGCACCTGGGCGCTCATGGCTTC 3856
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 2
US-09-106-182-2
; Sequence 2, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiostrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09106182

FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/051,053
FILING DATE: 30-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF385
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-182-2

Alignment Scores:
Pred. No.: 6,41e-57 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 1 Gaps: 2

US-09-931-704-3 (1-5087) x US-09-106-182-2 (1-225)

QY 3311 GGTGAGCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACTGGGCCCCCTTTCAACGAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCCCTCCCGCTGGGGGCGAGAGACTCTGCCAGGCCCACTGTGTGACTTG 3454
Db 72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGCGAAGCTCAATGACAACTGGCGTGAACCCAGAACTACGAGGCTACAGC 3514
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131
QY 3575 AGCTTGGCCCACTTCTGACACAGCTTCCAGAGGCTCTGGGCGAGCAATGCGGCGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTTGGGTACCCACTGCCCGAGCGCTGCCCTGGGAGTCAACCCACTTGGACTCCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
QY 3695 GCGCTTGGCCACAGTACTTCTCCAGAGAGATGGACGACTTCTGGCTGTGAAGGAGCTG 3754
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTTGGCTGTGGCGCTCGGCCAAGAGCTTCAACCGCTCAAGAGAGATGCAAGCCT 3814
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCAGTACCTGCACCTGGGCGCTCATGGCTTC 3856
Db 212 ProAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 3
US-08-988-819-2
; Sequence 2, Application US/08988819

```
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-988-819-2

Alignment Scores:
Pred. No.: 6.41e-57 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 3 Gaps: 2

US-09-931-704-3 (1-5087) x US-08-988-819-2 (1-225)
QY 3311 GGTGAGCCAGCGCTGCTGCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGCGCCCTTTCACAGCAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCAACCCCTCCCGCTGGGGCAGAGACTCTGCCAGGGCCACTGTGACTTG 3454
Db 72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGCGAAGCTCAATGACAACTCCGGCTGACCCAGAACTACGAGGCTACAGC 3514
Db 92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTGTGCTGCTGGCCCTCAACCGTCAGSCTGCCACTGTGCTGCGCGC 3574
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTGTGGCCACTTCTGCACAGCCCTCCAGGGCTCTGCTGGGCGAGCATTTGGGGCTCATG 3634
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCCACTGCGCCCGAGCCGCTGCTGGGAGCTGAACCCACTGTGACTCT 3694
Db 152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171

; 3695 GGCCTGCCACAGTGAATCTCTCCAGAGATGGACGACTTCTGCTGCTGAAGAGCTG 3754
; 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeu 191
; 3755 CAGACCTGGCTGTGGCGCTCGGCCAAGGACTTCAACCGCTCAAGAGAGATGACGCT 3814
; 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
; 3815 CCAGCAGTGCAGTCACCTGCACCTGGGGCTCATGGCTTC 3856
; 212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 4
US-09-016-534-2
; Sequence 2, Application US/09016534
; Patent No. 6143874
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; APPLICANT: ELLIOTT, GARY S.
; APPLICANT: SARMIENTO, ULIA
; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,534
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442B
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-016-534-2

Alignment Scores:
Pred. No.: 6.41e-57 Length: 225
Score: 896.00 Matches: 170
Percent Similarity: 90.72% Conservative: 6
Best Local Similarity: 87.63% Mismatches: 6
Query Match: 9.50% Indels: 12
DB: 4 Gaps: 2

US-09-931-704-3 (1-5087) x US-09-016-534-2 (1-225)
QY 3311 GGTGAGCCAGCGCTGCTGCCTCC-----TCCTTCCCATCACC----- 3349
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGCGCCCTTTCACAGCAG 3394
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
```

QY 3395 CCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGACTTG 3454
Db |||||
72 ProAspPheAsnProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeu 91
QY 3455 GAGGTGTGGGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTACAGC 3514
Db |||||
92 GluValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCGC 3574
Db |||||
112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131
QY 3575 AGCTGGCCCACTTCGACACCTCCAGGCTGCTGGGCGAGCAATTGGGGGCTCATG 3634
Db |||||
132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCACTGCCCGCTGGCTGGGAGTGAACCCACTTGGACTCCT 3694
Db |||||
152 AlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrPro 171
QY 3695 GGCCTTGGCCACAGTCACTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGGAGCTG 3754
Db |||||
172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGCTGGCGCTGGCCAAAGGACTTCAACCGGCTCAAGAAAGAGATGACGCT 3814
Db |||||
192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCAGTCACTCCCTGACCTGGGGGCTCATGGCTTC 3856
Db |||||
212 ProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 225

RESULT 5

US-08-792-019B-5
; Sequence 5, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-792-019B-5
Alignment* Scores:
Pred. No.: 1,55e-54 Length: 225
Score: 863.00 Matches: 163
Percent Similarity: 88.66% Conservative: 9
Best Local Similarity: 84.02% Mismatches: 10

Query Match: 9.15% Indels: 12
DB: 1 Gaps: 2
US-09-931-704-3 (1-5087) x US-08-792-019B-5 (1-225)
QY 3311 GGTGAGCAGCGGTCTGCTCCCTCC-----TCCTTCCCATCACC----- 3349
Db |||||
32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCCTTTTCACAGCTGAACCTACCTGGGCCCCCTTTCACAGG 3394
Db |||||
52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGlu 71
QY 3395 CCAGACTTCAACCTCCCGCTGGGGGAGAGACTCTGCCAGGGCCACTGTGACTTG 3454
Db |||||
72 ProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGTGGGAAGCTCAATGACAACTCGCGCTGACCCAGAACTACGAGGCTACAGC 3514
Db |||||
92 GluValTrpArgSerLeuAsnAspArgLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGGTGGCTCAACCGTCAGGCTGCCACTGCTGAGCTGCGCGC 3574
Db |||||
112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArg 131
QY 3575 AGCTTGGCCCACTTCTGACACCTCCAGGCTGCTGGGCGAGCAATTGGGGGCTCATG 3634
Db |||||
132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGTACCACTGCCCGCTGGCTGGGAGTGAACCCACTTGGACTCCT 3694
Db |||||
152 AlaThrLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProAlaTrpAlaPro 171
QY 3695 GGCCTTGGCCACAGTCACTTCTCCAGAGATGACGACTTCTGGCTGCTGAAGGAGCTG 3754
Db |||||
172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGCTGGCGCTGGCCAAAGGACTTCAACCGGCTCAAGAAAGAGATGACGCT 3814
Db |||||
192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCAGTCACTCCCTGACCTGGGGGCTCATGGCTTC 3856
Db |||||
212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225

RESULT 6

US-08-988-819-5
; Sequence 5, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/988,819
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/792,019
; FILING DATE: 03-FEB-1997
; ATTORNEY/AGENT INFORMATION:

NAME: COOK, ROBERT R.
 REGISTRATION NUMBER: 31,602
 REFERENCE/DOCKET NUMBER: A-442A
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 225 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-988-819-5

Alignment Scores:
 Pred. No.: 1,55e-54 Length: 225
 Score: 863.00 Matches: 163
 Percent Similarity: 88.66% Conservative: 9
 Best Local Similarity: 84.02% Mismatches: 10
 Query Match: 9.15% Indels: 12
 DB: 3 Gaps: 2

US-09-931-704-3 (1-5087) x US-08-988-819-5 (1-225)

```

QY 3311 GGTGAGCCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTCACAGCTGAACCTACCTGGGCCCCCTTTCAACGAG 3394
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCACACCTCCCGCTGGGGGAGAGAGACTGCGCCAGGGCCACTGGTGTGACTG 3454
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 ProAspPheAsnProArgLeuGlyAlaGlyThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGTGGGAAGCTCAATGACAACTCGGCTGACCCAGAGACTACGAGGCTACAGC 3514
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 GluValTrpArgSerLeuAsnAspArgLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGGCTGGCTCAACCGCTGCGGCTGACCTGCTGAGCTGCGCGC 3574
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTTGGCCCACTTCTGCACAGCCCTCCAGGGCTGCTGGGAGGCTGCTGCTGCTGCTG 3634
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGCTACCACTGCCAGCCGCTGCTGGGACTGAACCCACTTGGACTCCT 3694
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 AlaThrLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProAlaTrpAlaPro 171
QY 3695 GGCCTTGGCCACAGTACTTCTCCAGAGATGAGACGACTTCTGGCTGCTGAAGAGCTG 3754
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGCTGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 3814
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
QY 3815 CCAGAGCTGCAGTACCCCTGACCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 3874
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225

```

RESULT 7

US-09-016-534-5
 Sequence 5, Application US/09016534
 Patent No. 6143874
 GENERAL INFORMATION:
 APPLICANT: CHANG, MING-SHI
 APPLICANT: ELLIOTT, GARY S.
 APPLICANT: SARMIENTO, ULLA
 APPLICANT: SENALDI, GIORGIO
 TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:

ADDRESSEE: AMGEN INC.
 STREET: ONE AMGEN CENTER
 CITY: THOUSAND OAKS
 STATE: CA
 COUNTRY: USA
 ZIP: 91320
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/016,534
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/792,019
 FILING DATE: 03-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: COOK, ROBERT R.
 REGISTRATION NUMBER: 31,602
 REFERENCE/DOCKET NUMBER: A-442B
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 225 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-016-534-5

Alignment Scores:
 Pred. No.: 1,55e-54 Length: 225
 Score: 863.00 Matches: 163
 Percent Similarity: 88.66% Conservative: 9
 Best Local Similarity: 84.02% Mismatches: 10
 Query Match: 9.15% Indels: 12
 DB: 4 Gaps: 2

US-09-931-704-3 (1-5087) x US-09-016-534-5 (1-225)

```

QY 3311 GGTGAGCCAGCGGCTCTGCCCTCC-----TCCTTCCCATCACC----- 3349
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32 GlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGlu 51
QY 3350 -----CTCTCTTTTCACAGCTGAACCTACCTGGGCCCCCTTTCAACGAG 3394
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 HisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProPheAsnGlu 71
QY 3395 CCAGACTTCACACCTCCCGCTGGGGGAGAGAGACTTCTGCCAGGGCCACTGTTGACTG 3454
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 ProAspPheAsnProArgLeuGlyAlaGlyThrLeuProArgAlaThrValAsnLeu 91
QY 3455 GAGGTGTGGGAAGCTCAATGACAACTGCGGCTGACCCAGAGACTACGAGGCTACAGC 3514
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 GluValTrpArgSerLeuAsnAspArgLeuArgLeuThrGlnAsnTyrGluAlaTyrSer 111
QY 3515 CACCTTCTGTGTACTTGGCTGGCTCAACCGCTGCGGCTGACCTGCTGAGCTGCGCGC 3574
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 HisLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaThrAlaGluLeuArgArg 131
QY 3575 AGCTTGGCCCACTTCTGCACAGCCCTCCAGGGCTGCTGGGAGGCTGCTGCTGCTGCTG 3634
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 SerLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMet 151
QY 3635 GCAGCTCTGGGCTACCACTGCCAGCCGCTGCTGGGACTGAACCCACTTGGACTCCT 3694
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 AlaThrLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProAlaTrpAlaPro 171
QY 3695 GGCCTTGGCCACAGTACTTCTCCAGAGATGAGACGACTTCTGGCTGCTGAAGAGCTG 3754
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 GlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeuLysGluLeu 191
QY 3755 CAGACTGTGGCTGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 3814
|||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 192 GlnThrTyrLeuTipArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnPro 211
Qy 3815 CCAGAGCTGCACTACCTGACCTGGGGCTCATGGCTTC 3856
Db 212 ProAlaAlaSerValThrLeuHisLeuGluAlaHisGlyPhe 225

RESULT 8

US-08-963-825-21

; Sequence 21, Application US/08963825

; Patent No. 6110689

; GENERAL INFORMATION:

; APPLICANT: Ovisit, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/963,825

; FILING DATE:

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/187,319

; FILING DATE: 21-JAN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adda C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1078 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN ALPHA 1 (III)

US-08-963-825-21

Alignment Scores:

Pred. No.:

Score:

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

7.69e-16

330.50

29.53%

23.49%

3.54%

3

1078

319

82

502

456

71

US-09-931-704-3 (1-5087) x US-08-963-825-21 (1-1078)

Qy 5059, GGGAAACCTGGGTGGACAGAGTGTGTGGAGGACTGAGGGATCAAGGGTGTGGT 5000

Db 69 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84

Qy 4999 GGCAGCAGCAGTGGTAACTGAGTGTGCGCCATCGCCCGAGTGGCGGGGAGGAGGG 4940

Db 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96
Qy 4939 ---CCTCTTCCAGAGCTCTGTGCTGCCACACAGAGCAAGCAGGGAGTGT----- 4892
Db 97 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112
Qy 4891 ---GGAGCCAGGGGTGGCAAGAGTCTGATGAGCACCAGAGGGGTGAGCTGGGCAAGA 4835
Db 113 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu--- 130
Qy 4834 GACTAAGTCTGGGGCAAGCATTTTATTGTTAATACAGATAGAAATTCGCAATAAA 4775
Db 130 ----- 130
Qy 4774 TATCATCTAATAATAACATCTCCAAATAATAATAATAATAATAACACACTTAGAGTCA 4715
Db 130 ----- 130
Qy 4714 TGAGTGGGTGGGGCTGGGGGCGAGGCGCTTGGGGAGCTGCCA----- 4673
Db 131 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 149
Qy 4672 ---CCCTACACCCCAAAATGCTACTGCAATTGTAACCTTCAGGAAATCCTGTGGTGTGC 4616
Db 150 GlyProMetGlyProArg----- 155
Qy 4615 TATGGTTGGCCTCCCGAGCCTGGCAACCCACAGATACCTTGGGAAAGGGGCGAGAGA 4556
Db 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 173
Qy 4555 GGCACCGTAACTACTCGAAGACAATTCGAGGCAAGTCTGATGCTCAGCTCGGTTTGT 4496
Db 174 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 188
Qy 4495 TTGCCACTCTGTCTTTGGAGACTCTCGGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 4436
Db 189 GlyProProGlyThrAlaGlyPheProGly----- 198
Qy 4435 TTCAGTATTTCTCTCTCTGTTTATAGGGCACCTCCAATGCAGACCCCAACCTGCAC 4376
Db 199 -----SerProGlyAlaLysGlyGluAlaGlyProAlaGlySerPro----- 212
Qy 4375 TTAATCTCATCTCTTTTGTGTAGAAATTCGCCACCAATTTGTTTCTGTAATTTGGATGTAT 4316
Db 212 ----- 212
Qy 4315 AGAGTATTTGGGATAGGAGGAGCAACCCATCCAAAGAGAGTGGGGGAGCAAGGGC 4256
Db 213 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 224
Qy 4255 TTCACCCAGAAAGGAGTGTGTGGTGGCGCCCTCAACCCCTGCCCTCCAGATGTGCCA 4196
Db 225 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 237
Qy 4195 CCTGAGGGGCTGGGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4136
Db 238 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 252
Qy 4135 CTCTGCACCAACCTGAACCACTTCACTCCCTCAGCATGACTTCTCTAGTAAACAGGA 4076
Db 253 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 264
Qy 4075 CAGGACAGGGCTACTGTACT 4022
Db 265 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 284
Qy 4021 GGAGGAGCAGGGTGTGTCATCATCACGCCCGAGGCTCCAGGAAAGGGCCAGAGCTC 3962
Db 284 aGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGlyG 304
Qy 3961 ACAGTCTCTGTCT-----CCTGGCTCAACAGTGT---GGCATACAGGGCTGGCTCTC 3911

304 uAlaGlyIleProGlyValProGlyAlaLysGlyLysGlyLysAspGly-SerProG 324

3910 ACAAGTGGGAGCAGCGGTTTTGAAGGGGGAGCCGAAGAGAGAGTCTCAGAAGTCAGAAGCC 3851

324 LysAspProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaLeuGlySerA 344

3850 ATGAGCCCCAGCTGCAGGGTGACTGCTGCTGGAGGTGCATCTCTCTCTGTGAGCCG 3791

344 rgGlyProAlaGly----- 348

3790 GTTGAAGTCTCTGGCCGAGCCACAGCCAGGTCTGCAGCTCTCTTACGACGCCAGAGTC 3731

349 -----ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluArgGlyA 364

3730 GTCCATCTTCTGAGGAAGTCACTGTGGGCAGCGCCAGGAGTCCAAGTGGTTCAGTCCC 3671

364 laProGlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArg----- 379

3670 AGGCACGGCTGGGCAGTGGTAGCCAGAGCTGCCATGACGCCCGCAATGTGCCAG 3611

380 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---SerProG 396

3610 CAGGCCCTGGAGCTGTGCAGAAGTGGGCCAGGCTCGCGCGCAGCTCACGACGTGGCAGC 3551

396 lGlyProGlySerAspGlyLysProGlyPro----- 406

3550 CTGACGGTTGAGGCCACGCAAGTAAACACAGAAGTGGCTGTAGGCCTCGTAGTCTGGGT 3491

407 -----ProGlySerGlnGlyG 412

3490 CAGCCCGATTTGCTATTGAGGTTCCGCACACTCCAAGTCAACAGTGCCTCGTGGCAG 3431

412 lueR-----GlyArgProGlyProProGlyProSerGlyProArgGlyG 427

3430 AGTCTCTGCCCCAGCGGGAGGTTGAAGTCTGCTCGTTGAAAGGGGGCCCAAGTTA 3371

427 In-----ProGlyValMetGly-----PheProGlyProLysGlyAsnAspGlyA 442

3370 GTTCAGCTGTGAAAAGGAGAGAGGTGATGGGAAGAGGAGGAGCGCTGGCTCACC 3311

442 laProGlyLys-AsnGlyGluArgGlyGlyProGlyGly----- 454

3310 ACCAAGGATACCTGCTCCCAAAGTTCTATTTTTTGTGTCCCTGCACACTGGCCCTGT 3251

455 -----ProGlyProGlnGlyPro-----ProGly 462

3250 CTCCATGCTAGGCTCTTTGTTTCATCGCTCCCTGCTGAGTTTTTCCAATAGGATATCATTT 3191

463 LysAsnGlyGluTyGlyProGlnGlyProPro----- 473

3190 GTATGCAAGTGCAGTTGAAGCAGTGGCGGTGGTGAGCTCTCCAATGGAGAGCGGTAGAA 3131

474 -----GlyPro 475

3130 AAAGAGGAGGAGGAGACGCAACAAGAC---GCTCAGGAAGGAGCTGCTGATTGGCTGGGT 3074

476 ThrGlyProGlyGlyAspLysGlyAspThrGlyProArgGlyProGlnGlyLeuGlnGly 495

3073 GGGCAACCTTGGGAGATGCCCGCTGTGGGGCAGGACAGAGAGAAAGAGAGCCTGGA 3014

496 LeuProGlyThrGlyGlyProPro-----GlyGluAsnGluLysProGlyGluProGly 513

3013 GGAGCTGGAGAGTGAAGGTGGCCAGGAGAGTGCAGAGACCAAGAAAGAGAGGTTCCA 2954

514 ProLysGlyGluAla---GlyAlaProGlyAlaProGlyGlyGlyAspAlaGlyAla 532

2953 GAAGGAGCAGCGCTCAGCAGTGTGGCTGCCAAAACAGACAGCAAGGAGGAATGAGCCTTCA 2894

533 ProGly-----GluArgGlyProPro 539

2893 AAACACTGTCGGGCT-----TGGCAGAGAGGTGCTGACACTTGGAGAGAACCATTTG 2840

540 GlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProPro----- 554

QY	2839	CTGGGGCTGGAAAGCCAGATTCGAGGGGGTTAAGAAGTGAAGTGGTGGAGGAAGTGGGA	2789
Db	555	---GlyPro-----GluGlyGlyLysGlyAlaAlaGlyProProGlyProPro	569
QY	2779	GGCAGAGGGTGTAGACACTCTCTTGGAGAAATTTGCAGGGAAGCCAGGAGAGAA---	2723
Db	570	GlyAlaAlaGlyThrProGlyLeu-----GlnGlyMetProGlyGluArgGly	585
QY	2722	-----AAGGTCGGGAACTAGAGAGGGCAGTGGAGTGAAGCAAA	2684
Db	586	GlyLeuGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAsp	605
QY	2683	GGTTTTCTGCACAAGGAACAGAGAGAGATGAGATGGCCCATTTGGAACACAGAGCCCCAT	2624
Db	606	GlyValProGlyLysAspGlyProArgGlyProThrGlyProIleGlyProGly---	624
QY	2623	TTGCCGCTGACACACCCAGGCCACACGCTGAGGGCTGGGAGAGACGGGTACTCACTA	2564
Db	625	---ProAlaGlyGlnProGlyAspLysGlyGluGly---GlyAlaProGlyLeuProGly	642
QY	2563	ACACTCGGAAGNAGGGTTCA---GCTGATAGGACAGGGTT-----	2525
Db	643	IleAlaGlyProArgGlySerProGlyGluArgGlyGluThrGlyProGlyProAla	662
QY	2524	-----GCAGAAGGAAMAATCGAGGCAGCTCATGATAGGACAGGAAAGG	2477
Db	663	GlyPheProGlyAlaProGlyGlnAsnGly-----GluProGlyGly---	676
QY	2476	CTATGGAAGACAGCATGGTATAGGGAAGGGCAGAGCCCTTTCCAAACCCTCGGGGAGC	2417
Db	677	-----LysGlyGlu-----	679
QY	2416	GGCTGGGGCCAGGAAAGGGATGAGGGAGATCTATAACTGGGAAGAGAGGGAAGCAAA	2357
Db	680	ArgGlyAlaProGlyGlu-----LysGlyGlu	698
QY	2356	GGGAGAGCCCTGGGATAGGCAAGGCCCGAGGGAGGTTTCAACGCTCAACAACACAGA	2297
Db	689	GlyGlyProGlyValAlaValProProGlyGlySerGlyProAlaGly-ProProGly	708
QY	2296	GACAACTAATTTCTCGGTGTCTCTGGGCTGGAGCCAGCTCCCCGCCACCTCTCTC	2237
Db	708	y-----ProGlnGlyValLysGlyGluArgGlySerProGlyGlyProGlyAla	724
QY	2236	CCTTTTCTTCTGCTCCAGAGTAGGCTCGGATGTGGAAGAGAGAAACACGGGAAGAG	2177
Db	724	AlaGlyPhePro-----GlyAlaArg	731
QY	2176	GGGG---CGGTTTCCCATGTGTGCTCGGCTCCTCAACCCTCTCTCTCTCTCTCTGGG	2120
Db	731	gGlyLeuProGlyProProGlySerAsnGlyAsnProGly---ProProGlyProSerGly	750
QY	2119	CTCTCTGAACACCTCTGGACACCCCAAAATCCTTACAGCATAACTCCATCCACACCT	2060
Db	750	ySerProGlyLysAspGly---ProProGlyProAlaGlyAsnThrGly-----	765
QY	2059	AAACCCCTGCCACTACTCTGGTTCCTACGTCCCTTTCCCCCATGACCCCCCTAG	2000
Db	766	-----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaGly	780
QY	1999	GAGGCCCAGCTAGGATGGGAGGAGCAGCGGAAGGAGGCCCTCGGTCCAGAGCAGCAG	1940
Db	780	lyGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGly-----AlaProGly	798
QY	1939	GGCCC-----CAGGTCTGGGGCAGAGCGAGCTGGAAACGCGGCATGGGAGG	1892
Db	798	lyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP	818
QY	1891	TGGGGCCCCCTC---TGTCCCAACCAACCCCGCCGCGGAGAGACACCAACACGGGTG	1835
Db	818	roGlyProArgGlySerProGlyProGlnGlyValLysGlyGlyLysGlyProGlyAla	838

```

Qy 1834 TAAAC-----AGAGGAGACTTCCCGCGCCCGCCAGGGGCGCGGC-----1794
Db 838 laasnglyleuSerGlyGluAArgGlyPro-----ProGlyProGlnGlyLeuProGlyLeuA 857
Qy 1793 --GGTGGGCTGGACCCAGTGGGGCAATGGGCAAGGGGGTGGCAGGCTCCCGCAT 1736
Db 857 laGlyThrAlaGlyGluProGlyAArgAspGlyAsnProGlySerAspGlyLeuPro-----875
Qy 1735 GCCTAGTCTGGGGCCAGGATGGGAAGTGGGGGGCCACCTAAGAGGGAAGGCGAGGA 1676
Db 876 -----GlyArg-----AspGlySerProGlyGlyGlyAspA 887
Qy 1675 GATAGACACAGAGGGGAGTGGTGGGAGCCAAAGAGCCCTCTCAACCCCATCATTTAC 1616
Db 887 rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG 903
Qy 1615 CAGAACCTCCACGGTTAGACTGTCTTCCCAACTCTCACTCTCTCCCAACTCCAG 1556
Db 903 lyProProGly-----ProValGlyProA 911
Qy 1555 ATTCTACCTGGTACTACATAGTCCAGCCAGCTCGGAGTGGTGGTCCAGGTA 1496
Db 911 laGlySerGlyAspArgGlyGluSerGlyProAlaGlyProAlaGlyAlaProGlyP 931
Qy 1495 GCGGGTGGGTATAGTGTCTTGGATGGAGGG-----CCAGGCCCTGGGTCCCTGTGC 1440
Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgG 945
Qy 1439 GATTGAGCTGCATCGAGGAGGTGCCAGAGCACCCTGTCAGCCAGCAGCTTAACATCC 1380
Db 945 lyAspGlyGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGlyPheP 965
Qy 1379 CC-----CACGAGCTCTGTGGCGAGGATGGACGAGAAGCATG 1341
Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleLys 985
Qy 1340 AGCGCGCGCCCG-----GGCCACAGCGGGAAGCAGCTGCGAGCCCTCAGGTGTCTGTCCCC 1284
Db 985 erProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProGlyL 1003
Qy 1283 ACCTTGGCTGGGTGGGGGTGAGGGT-----GGGACACAGATGGGAGGCGAGGGGA 1230
Db 1003 yeAspGlyThrSer-GlyHisProGlyProIleGlyProGlyProArgGlyAsnArg 1022
Qy 1229 TGGAGAGGGGAAC-----ACTCTGGCAAGACGGAGAG 1194
Db 1023 GlyGluArgGlySerGluGlySerProGlyHisProGlyGln 1036

RESULT 9
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-500-811-21

Alignment Scores:
Pred. No.: 7,69e-16 Length: 1078
Score: 330.50 Matches: 319
Percent Similarity: 29.53% Conservative: 82
Best Local Similarity: 23.49% Mismatches: 502
Query Match: 3.54% Indels: 456
DB: 4 Gaps: 71

US-09-931-704-3 (1-5087) x US-09-500-811-21 (1-1078)
Qy 5059 GGGAAACCTGGGGTGGAGCAGAGCTGTGGGAGACTGAGGATCAAGGGGTGTGTGTT 5000
Db 69 GlyProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84
Qy 4999 GGCAGCAGCAGTGGGTAACTCTGAGGTGCGCCCATCGCCCGAGTGGCGGAGGAGAGGG 4940
Db 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96
Qy 4939 ---CCTCTTCCCAGAGCTCTGTGCAGCTCCACAGGCAAGCAGGAGTGTG-----4892
Db 97 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112
Qy 4891 ---GGAGCCAGGGTGGCAAGAGTCTGTAGCACCAGGAGGCTCAGGCTGGGCAAGA 4835
Db 113 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGluLysGlyGlu---130
Qy 4834 GACTAAGTCTGGGGCAAGCATTTTATTGTTTAAATACAAGATAGAAATTTCTGCAATAAA 4775
Db 130 -----130
Qy 4774 TATCATCTAATAATAACATCTCCAAATAATAATAATTAATACACACACTTAGATCA 4715
Db 130 -----130
Qy 4714 TGAGTGGTGGGGCTGGGGGGCAGGCGCTTGGGAGCTGCCA-----4673
Db 131 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 149
Qy 4672 ---CCCTACACCCCAAAATGCTACTGCTATTGTTAAACTTTTCAGGAATCTCTGTGTGGC 4616
Db 150 GlyProMetGlyProArg-----155
Qy 4615 TATGTTTGCCCTCCCGCAGCTCGCAACCCACAGATACCTCGGAAAGGGGCGAGAGA 4556
Db 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg---173
Qy 4555 GGCACCGTAATACTGGAAGACAATTCGAGGCAAGGTCCTGATCTCAGCTCGGGTTTGT 4496

```



```
QY 2416 GCGTGGGCCCCAGGAAAGGATGAGGAGATGTATATACTGGGAAAGAGGAGCA 2357
    |||||:::
Db 680 ArgGlyAlaProGlyGlu-----LysGlyGlu 688
    |||||:::

QY 2356 GGGAGAGGCTGGATAGCAAGGCCCCAGGGGAGGTTTCAACGCTGACAAGCACAGA 2297
    |||||:::
Db 689 GlyGlyProProGlyValAlaValProProGlyGlySerGlyProAlaGly-ProProG 708
    |||||:::

QY 2296 GACAACTAATTTCTCGGTATCTCTGGGCTGGAGCCAGCTCCCGCCACCTCTTTC 2237
    |||||:::
Db 708 y-----ProGlnGlyValysGlyGluArgGlySerProGlyGlyProGlyAl 724
    |||||:::

QY 2236 CTTTTTCTTCTCCAGTCCAGTAGGCTGTGGATGTGAAGAGAAACACCGGAAGAG 2177
    |||||:::
Db 724 aAlaGlyPhePro-----GlyAlaAr 731

QY 2176 GGGG---CGGTTCCCACTGGTCTCGGCTCTCAACCTCTCTCTCTCTCTCTCTGGG 2120
    |||||:::
Db 731 gGlyLeuProGlyProProGlySerAsnGlyAsnProGly---ProProGlyProSerG 750
    |||||:::

QY 2119 CTCTCTGAACACTCTGACACCCCAATCTTACAGCATAACTCCATCCACACCCCT 2060
    |||||:::
Db 750 ySerProGlyLysAspGly---ProProGlyProAlaGlyAsnThrGly----- 765

QY 2059 AAACCCCTGCCACTACTGCTGGGTTCCCTACAGTCCCTTCCCTCCATGACCCCTAG 2000
    |||||:::
Db 766 -----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaG 780

QY 1999 GAGGCCCACTAGTAGGAGGAGCAGAGGAAGGAGGCGCCCTGGTCCAGAGCAGCAG 1940
    |||||:::
Db 780 lGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGly-----AlaProG 798

QY 1939 GGGCC-----CAGTCTGGGGCAGAGCGGAGCTGGAACGACGCGCATGGGAGG 1892
    |||||:::
Db 798 lProLeuGlylleAlaGlylleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP 818

QY 1891 TGGGGCCCCCTC---TGTCCTCCACCAACCCCGCCGCGGAGAGACACCAACACCGGTG 1835
    |||||:::
Db 818 roGlyProArgGlySerProGlyProGlnGlyValysGlyGlySerGlyLysProGlyA 838

QY 1834 TAAAC-----AGAGGAGACTTCCCGGCCCCAGAGGGGAGCGCGGC----- 1794
    |||||:::
Db 838 laAsnGlyLeuSerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA 857

QY 1793 --GGCTGGGCTGCACCCAGTGGGGCAATGGGCCAAGCGGCTCGAGGCTGCCACAT 1736
    |||||:::
Db 857 laGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro---- 875

QY 1735 GCTAGTCTCGGGCCAGGATGGGAAGTGGGGGGGCGCCACCTAAGAGGGAAGGCGCAGGA 1676
    |||||:::
Db 876 -----GlyArg---AspGlySerProGlyGlyLysGlyAspA 887

QY 1675 GATAGACACAGAGGGGAGTGTGGGAGCCAAAGAGCCCTCTCACTCCCATCATTTAC 1616
    |||||:::
Db 887 rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG 903

QY 1615 CAGAACCTCCAGGTTAGACTGTCTTCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCC 1556
    |||||:::
Db 903 lProProGly-----ProValGlyProA 911

QY 1555 ATTCTACGCTGATCTACATAGTTCAGCCAGCTCGGAGTGTGGTGTCCAGGTA 1496
    |||||:::
Db 911 laGlyLysSerGlyAspArgGlyGlySerGlyProAlaGlyProAlaGlyAlaProGlyP 931

QY 1495 GCGGGTGAGGTATAGTTTTCTTGATGGAGGGG-----CCAGGCCCTGGGTCCCTGTGC 1440
    |||||:::
Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgG 945

QY 1439 GATTGAGAGCTGCACCTGCAGGGAGGTGCCAGACCGGTGCACAGGCAGCCTAACATCC 1380
    |||||:::
Db 945 lAspLysGlyGluThrGlyGluArgGlyAlaAlaGlylleLysGlyHisArgGlyPheP 965
```

```
QY 1379 CC-----CACAGTCCCTGTGGGCGAGGATGGACGAGAGCATG 1341
    |||||:::
Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlys 985
    |||||:::

QY 1340 AGCGGCGCCCG---GGCCACACGCGGAAGCAGCTGGCAGCCCCCTCAGGTCTCTCTCCC 1284
    |||||:::
Db 985 erProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyL 1003
    |||||:::

QY 1283 ACCTTGGGCTGGGTGGGGGTGAGGGT-----GGGACACCATGATGGGAGGCGAGGGA 1230
    |||||:::
Db 1003 ysAspGlyThrSer-GlyHisProGlyProIleGlyProProGlyProArgGlyAsnArg 1022
    |||||:::

QY 1229 TGAAGGAGGGAAC-----ACTCTGGCAAGACGGGAGAG 1194
    |||||:::
Db 1023 GlyGluArgGlySerGlyGlySerProGlyHisProGlyGln 1036
    |||||:::

RESULT 10
US-09-570-573-21
; Sequence 21, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Oviast, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-570-573-21

Alignment Scores:
Pred. No.: 7.69e-16 Length: 1078
Score: 330.50 Matches: 319
Percent Similarity: 29.53% Conservative: 82
Best Local Similarity: 23.49% Mismatches: 502
Query Match: 3.54% Indels: 456
```

DB: 4 Gaps: 71
US-09-931-704-3 (1-5087) x US-09-570-573-21 (1-1078)
QY 5059 GGAACCTGGGGTGGACAGAGCTGTGGGAGGACTGAGGATCAAGGGTTGGTGT 5000
Db 69 GlyProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84
QY 4999 GGCAGCAGCAGTGGTAACTCTGAGTGGCCCATCGCCAGTGGCGGAGGAAGAGG 4940
Db 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96
QY 4939 ----CTCTTCCAGAGCTCTGTGCAGCTCCACAGAGGCAAGCAGGAGTG----- 4892
Db 97 LeuProGlyProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112
QY 4891 ---GGAGCCAGGGTGGCAAGAGCTCTGATCAGCACAGAGAGGGTGAAGTGGCAAGA 4835
Db 113 ProGlyMetLysGlyHisArgGly-----PheAspGlyArgAsnGlyGluysGlyGlu--- 130
QY 4834 GACTAAGTTCTGGGCAAGCATTTTATTGTTAATACAGATAGAAATTTCTGCAATAA 4775
Db 130 ----- 130
QY 4774 TATCATCTTAATAATAACATCTCCAATAATAATAATAATAATAACACACTTAGAGTCA 4715
Db 130 ----- 130
QY 4714 TGAGTGGTGGGCTGGGGGAGGAGGCTTGGGGAGCTGCCA----- 4673
Db 131 ---ThrGlyAlaProGlyLysGlyGluysGlyLysGlyLeuProGlyGluAsnGlyAlaPro 149
QY 4672 ---CCCTACACCCAAATGCTACTGCTGTTAACTTTTCAGGAAATCTGTGTGTGGC 4616
Db 150 GlyProMetGlyProArg----- 155
QY 4615 TATGTTGCTCCCGAGCTGGCAACCCACAGATACCTCTGGGAAAGGGGAGAGAGA 4556
Db 156 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 173
QY 4555 GGCACGTAATACTGAAAGCAATTCGAGCAAGTCTGATGCTCAGCTCGGGTTTGT 4496
Db 174 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 188
QY 4495 TTGCCACTCTGCTTTGGAGACTCTCTGGGTCTGTCTGTCTCTCCCTGCCCTGTGCT 4436
Db 189 GlyProGlyThrAlaGlyPheProGly----- 198
QY 4435 TTCAATTTCTCTCTCTGTTATAGGCACCTCCAAATGCAGACCCCAACCTGCACTG 4376
Db 199 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro----- 212
QY 4375 TTAATCTCATCTCTTTTGTGTAGAATTGCCACCATGTTTGTCTCTGAATGATGAT 4316
Db 212 ----- 212
QY 4315 AGAGTGATTTGGGATAAGGAGCAACACCCATCCAGAGAGAGTGGGGAGCCAAAGGGC 4256
Db 213 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 224
QY 4255 TTCACCCAGAAAGGCATGTGTGTGGTGGCCCTCAACCCCTGCCCTCCAGATGTGCCA 4196
Db 225 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 237
QY 4195 CCTGAGGGCTGGTGGGCACTGCCAGTGGTAGGCAAGAGAGAGGAGGCGCATGAGCGC 4136
Db 238 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 252
QY 4135 CTCTGCACCAACCTGAACCACTTCACACTCCCTCGACAGTACTTCTCTGTAGAAACAGGA 4076
Db 253 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 264
QY 4075 CAGGACAGGGCCCTACTGTACTCTCTCCAGCTCGGTAGACCT-----TTGGGAGGTGG 4022

Db 265 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 284
QY 4021 GGAGAGACAGGCTGTATCGCATCACAGCCCGGTCCAGGAAAGGGCCAGAGGCTC 3962
Db 284 aGlyGluProGlyLysAsnGlyAlaLysGlyGluProGlyProArgGlyGluArgGlyG 304
QY 3961 ACAGCTTCTGTCT-----CTGGCTCAACAGGTGTT---GGCATACAGGCTGGCTCTC 3911
Db 304 uAlaGlyIleProGlyValProGlyAlaLysGlyGluAspGlyLysAspGly--SerProG 324
QY 3910 ACAAGTGGGAGCAGGTTTGAAGGGGAGCAAGAGGAGAGTCCAGAGTCAGAAAGCC 3851
Db 324 lyAspProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyAlaLeuGlySerA 344
QY 3850 ATGAGCCCCAGGTGCAGGCTGACTGCAGCTCTGGAGGCTGCATCTTCTTGTAGCCG 3791
Db 344 rgGlyProAlaGly----- 348
QY 3790 GTTGAAGTCTTGGCCGAGCCACAGCAGGTCTGCAGTCTCTTACAGCAGCAGAGTCC 3731
Db 349 -----ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluArgGlyA 364
QY 3730 GTCCATCTTGGAGGAAGTCACTGTGGGAGGCGCAGGAGTCCAGAGTGGGTTCAGTCCC 3671
Db 364 laProGlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArg----- 379
QY 3670 AGCAGCGGCTTGGGAGTGTAGCTGAGCTGCCAGCTCCATGACGCCCAATGCTGCCAG 3611
Db 380 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---SerProG 396
QY 3610 CAGCCCTGGAGGCTGTGCAGAGTGGGCCAGGCTGCGGCGCAGCTCAGCAGTGGCAGC 3551
Db 396 lyGlyProGlySerAspGlyLysProGlyPro----- 406
QY 3550 CTGAGGTTGAGGCCACCAAGTAACAGAGGTGGTGTAGGCTCTAGTCTGGGT 3491
Db 407 -----ProGlySerGlnGlyG 412
QY 3490 CAGCGCAGTTTGTATTGAGGCTTTCGCACACCTCCAAAGTCAACAGTGGCCCTGGGCG 3431
Db 412 luSer-----GlyArgProGlyProProGlyProSerGlyProArgGlyG 427
QY 3430 AGTCTGCCCCCAGCGGGGAGGTTGAAGTCTGGCTCGTTGAAAGGGGGGCCAGGTA 3371
Db 427 ln-----ProGlyValMetGly-----PheProGlyProLysGlyAsnAspGlyA 442
QY 3370 GTTCACTGTGAAAGAGAGAGGTTGATGGGAAAGAGAGGAGGCGCCTGGCTCACC 3311
Db 442 laProGlyLys--AsnGlyGluArgGlyGlyProGlyGly----- 454
QY 3310 ACCAAGGAGATACCTGCTCCCAAGTTTCTATTTTGTGTCCTCCCTGACACTGGCCCTGT 3251
Db 455 -----ProGlyProGlnGlyPro-----ProGly 462
QY 3250 CTCCATGTAGGCTCTTTTGTTCATGGCTCCCTGAGTGTTCCTCAATAAGGATATCATTT 3191
Db 463 LysAsnGlyGluTyrGlyProGlnGlyProPro----- 473
QY 3190 GTATGAGGTGACAGTTGAAGCAGTGGCGGTGGTGAGCTCTCAATGGAGAGCGGTAGAA 3131
Db 474 -----GlyPro 475
QY 3130 AAAGGAGGAGGAGAGCGCACAGAC---GCTCAGAAAGGAGCTGTGTTGGCTGGGT 3074
Db 476 ThrGlyProGlyGlyAspLysGlyAspThrGlyProArgGlyProGlnGlyLeuGlnGly 495
QY 3073 GGGCAACCTTTGGGAGAGTGGCCGCTGTGGGGGAGGAGAGAGAGAGAGAGGAGCTGGA 3014
Db 496 LeuProGlyThrGlyGlyProPro-----GlyGluAsnGluLysProGlyGluProGly 513
QY 3013 GGAGCTGGAGAGGTAGAAGGTGGCCAGAGAGTGCAGAGAGTGCAGAGAGAGGAGGTTCGA 2954

Db 514 ProLeuGlyGluAla---GlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAla 532
Qy 2953 GAAGGAGCAGCGCTCAGCAGTGTGGCTGCAAAACACAGACAGCAAGGAGATGACCTTCA 2894
Db 533 ProGly-----GluArgGlyProPro 539
Qy 2893 AAACACTGTGGGCT-----TGGCAAGAGGTGCTTGCAGACCTTGGAGAGAACCACTTG 2840
Db 540 GlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro----- 554
Qy 2839 CTGGGCTTGGAGCAGATGAGGGGGTTAAGAGTGTGGTGGAGAGGAAGTGGGA 2780
Db 555 ---GlyPro-----GluGlyGlyLysGlyAlaAlaGlyProProGlyProPro 569
Qy 2779 GGCAGAGGTGTAGACACTCTTGGAGAAATTGGCAGGGAAGCAGGAGAGAGAAA--- 2723
Db 570 GlyAlaAlaGlyThrProGlyLeu-----GlnGlyMetProGlyGluArgGly 585
Qy 2722 -----AAGGGTTCGGAACTAGAGAGGGCAGTGGAGTGAAGCAAA 2694
Db 586 GlyLeuGlySerProGlyProLysGlyAspLysGlyGluProGlyGlyProGlyAlaAsp 605
Qy 2683 GGTTCCTGCACAGAAACAGAGAGATGAGAGTGGCCCATTTGAAACACAGAGCCCAT 2624
Db 606 GlyValProGlyLysAspGlyProArgGlyProThrGlyProLleGlyProProGly--- 624
Qy 2623 TTGCGCTGACACACAGCCAGCCACAGCTGAGGGCTGGGAGAGACGGGTACTACTA 2564
Db 625 ---ProAlaGlyGlnProGlyAspLysGlyGluGly---GlyAlaProGlyLeuProGly 642
Qy 2563 ACACCTCCGAAGAAGGTTC---GCTGATAGGAGCAGGGTT----- 2525
Db 643 IleAlaGlyProArgGlySerProGlyGluArgGlyGluThrGlyProProGlyProAla 662
Qy 2524 -----CGAAGAAGAAATGAGAGCGAGTCTCATGATGAGACAGAGGAGG 2477
Db 663 GlyPheProGlyAlaProGlyGlnAsnGly-----GluProGlyGly--- 676
Qy 2476 CTATGAAGACACATGTTATAGGAAGAGGCGACAGCCCTTTCCAACTCCGGGGAGC 2417
Db 677 -----LysGlyGlu 679
Qy 2416 GGCTGGGGCCAGGAAAGGATGAGGAGATGTATACTGGGAAAGAGGAGCAAA 2357
Db 680 ArgGlyAlaProGlyGlu-----LysGlyGlu 688
Qy 2356 GGGGAGGCTGGATAGCAGAGGCGCCAGGGGAGGTTCACAGCTGACACACAGA 2297
Db 689 GlyGlyProProGlyValAlaValProProGlyLysSerGlyProAlaGly-ProProGly 708
Qy 2296 GACAACTAATTTCTCGGTGTATCTCGGGCTGGAGCCAGCTCCCGCCACCTTCTTC 2237
Db 708 y-----ProGlnGlyValLysGlyGluArgGlySerProGlyGlyProGlyVal 724
Qy 2236 CTTTCTTTCTTCCTGCTCCAGAGTAGGTCTGGGATGTGGAAGAGAACACCGGGAAGAG 2177
Db 724 aaAlaGlyPhePro-----GlyAlaAr 731
Qy 2176 GGGG---CCGGTTCCTCATGTGCTCGGTCTCTCCAACTCTTCTCTCTCTCTCTGGG 2120
Db 731 ggGlyLeuProGlyProProGlySerAsnGlyAsnProGly---ProProGlyProSerGly 750
Qy 2119 CTCTCTGAACACCTCTGGACACACCCCAATCTTACAGCATACCTCCATCCACACCT 2060
Db 750 ySerProGlyLysAspGly---ProProGlyProAlaGlyAsnThrGly----- 765
Qy 2059 AAACCCCTGCCACTACTGCTGGTTCCTTACAGTCCCTTCCCTCCATGACCCCTTAG 2000
Db 766 -----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaG 780
Qy 1999 GAGGCCAGCTAGATGGGAGAGCAGACGAGGAGGAGGCGCTGGTCCAGCAGCAG 1940
Db 780 yGlnProGlyGlyLysGlySerProGlyAlaGlnGlyProProGly-----AlaProG 798

Qy 1939 GGGC-----CAGGTCTGGGGCAGAGGCGAGCTGGAACCGCAGGCATGGGAGG 1892
Db 798 lyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP 818
Qy 1891 TGGGGCCCTC---TGTCCTCCACCAACCCCGCGCAGAGACACACACACACGCGTG 1835
Db 818 roGlyProArgGlySerProGlyProGlnGlyValLysGlyLysGlyLysProGlyA 838
Qy 1834 TAAAC-----AGAGGAGACTTCCCGGCCCCCAGGGGGAGGGCGGCG--- 1794
Db 838 laAsnGlyLeuSerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA 857
Qy 1793 --GGCTGGCTGGCAGCAGCGTGGGCAATGGGCCAAGCGCGGTGCGAGCGCTGCCACCAT 1736
Db 857 laGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeuPro--- 875
Qy 1735 GCCTAGTCTCTGGGCGCAGGATGGGAAAGTGGGGGGCCACCTAAGAGAGGAGGCGGAGA 1676
Db 876 -----GlyArg---AspGlySerProGlyGlyLysGlyAspA 887
Qy 1675 GATAGACCAGACAGGGGACTGGTGGGAGCCCAAGAGCCCTCTCTCACCCCATTTAC 1616
Db 887 rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG 903
Qy 1615 CAGAACCTCTCCACGGTTAGGACTGTCTTCCCAACTCTCTACTCTCTCCCAACTCCAG 1556
Db 903 lyProProGly-----ProValGlyProA 911
Qy 1555 ATTCTACCTGGTATCTACATAGTCCAGCCCAAGCTGCGGAGTTGTGTCTCCAGGTA 1496
Db 911 laGlyLysSerGlyAspArgGlyGluSerGlyProAlaGlyProAlaGlyAlaProGlyP 931
Qy 1495 GCGGTGAGGTCTAGTGTCTTCTGATGGAGGG---CCAGGCCCTGGGTCCCTGTGC 1440
Db 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgG 945
Qy 1439 GATTGAGAGCTGGCACTGAGGAGGTCCAGGAGCCGTCACAGCCAGCTTAACATCC 1380
Db 945 lyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGlyPheP 965
Qy 1379 CC-----CAGAGTCCCTGTGGGCGAGGTGGACGAGAGCATG 1341
Db 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlyS 985
Qy 1340 AGCGGGCCCG---GGCCACACCGGGAAGCAGCTGGCAGCCCTCAGGTGTCTGTCCCC 1284
Db 985 erProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyL 1003
Qy 1283 ACCTTGGCTGGTGGGGGTGAGGGT-----GGGACACACAGATGGGAGGAGAGGGA 1230
Db 1003 ysAspGlyThrSer-GlyHisProGlyProLleGlyProProGlyProArgGlyAsnArg 1022
Qy 1229 TGGAGGAGGGAAAC-----ACTCTGGCAAGACGGGAGAG 1194
Db 1023 GlyGluArgGlySerGluGlySerProGlyHisProGlyGln 1036

RESULT 11

US-09-548-608-21

; Sequence 21, Application US/09548608

; Patent No. 6355442

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

STATE: New York
 COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/548,608
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/187,319
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1078 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN ALPHA 1 (III)
 US-09-548-608-21

Alignment Scores:
 Pred. No.: 7,69e-16 Length: 1078
 Score: 330.50 Matches: 319
 Percent Similarity: 29.53% Conservative: 82
 Best Local Similarity: 23.49% Mismatches: 502
 Query Match: 3.54% Indels: 456
 DB: 4 Gaps: 71

US-09-931-704-3 (1-5087) x US-09-548-608-21 (1-1078)

QY 4672 ---CCCTACACCCCAAAATGCTACTGATTTGTAACCTTTTCAGGAATCCTGTGTGTGGC 4616
 DB 150 GlyProMetGlyProArg
 QY 4615 TATGTTGCCCTCCCGAGCTGGCAACCCACAGATACCTCTGGGAAAGGGGAGAGAAGA 4556
 DB 156 ---GlyAlaProGlyGluArgGlyProGlyLeuProGlyAlaAlaGlyAlaArg--- 173
 QY 4555 GGCACCGTAATACTGGAAGACAATTCGAGCAAGGCTCTGATGCTGCTGGGTTTGT 4496
 DB 174 -----GlyAsnAspGlyAlaAlaGlySerAspGlyGlnProGlyProPro 188
 QY 4495 TTGCCACTCTGTGCTTTGGAGACTCCTGGTCTGGTCTGCTCTGCTCCCTCCCTGTGCT 4436
 DB 189 GlyProProGlyThrAlaGlyPheProGly
 QY 4435 TTCAGTATTTCTCTCTGTTTATAGGCACCTCCAATGCAGACCCCAACCTTGCACGTG 4376
 DB 199 -----SerProGlyAlaGlyGluValGlyProAlaGlySerPro----- 212
 QY 4375 TTAATCTCATCTCTTTTGTGTAGATTGCCACCATGTTTGTCTTCTGAATTGATGTAT 4316
 DB 212 ----- 212
 QY 4315 AGAGTGATTGGGGATAAGGGAGCAACCCATCCCAAGAGAGTGGGGAGCAAAAGGC 4256
 DB 213 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 224
 QY 4255 TTCACCCAGAAAGCATGTGTGGTGGCCCTCAACCCCTGCCCTCCAGATGTGCCA 4196
 DB 225 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 237
 QY 4195 CCTGAGGGCTGGGTGGGCACTGGCCNAGTGTAGCCAGAGAGACAGAGCCATGAGCGC 4136
 DB 238 ProProGlyIleAsnGlySer---ProGlyGlyGlyGlyGlyMetGly----- 252
 QY 4135 CTCTGCACCAACCTGAACCATCTTCACATCCCTCCAGCATGACTTCTGTAGAAACAGGA 4076
 DB 253 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 264
 QY 4075 CAGGACAGGGCTACTGTACTCTCTCCCGAGCTCGGTAGACCT-----TTGGGAGGTGG 4022
 DB 265 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 284
 QY 4021 GGAGAGACAGGGCTGATCGCATCACGCCAGCCGGTCCAGAGAAAGGGCAGAGCTC 3962
 DB 284 aGlyGluProGlyGlyAsnGlyAlaGlyGlyGluProGlyProArgGlyGlyGlyGly 304
 QY 3961 ACAGTTCTGTCT-----CCTGGCTCAACAGGTGT---GGCATACAGGGCTGGCTCTC 3911
 DB 304 uAlaGlyIleProGlyValProGlyAlaGlyGlyGlyGlyAspGly-SerProG 324
 QY 3910 ACAAGTGGAGAGGGTTTGAAGGGGAGGAGAGAGAGAGAGGTGAGAGTCAAGTCAAGCC 3851
 DB 324 lyAspProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGlyGlyAlaLeuGlySerA 344
 QY 3850 ATGAGCCCCAGGTGCGAGGTGACTGCGAGCTGCTGGAGGCTGCATCTTCTTTGAGCCG 3791
 DB 344 rgGlyProAlaGly
 QY 3790 GTTGAAGTCTTGGCCGAGCCGCCAGCCAGGTCTGAGCTCTTTCAGCAGCCAGAGTCTC 3731
 DB 349 -----ProAsnGlyIleProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 364
 QY 3730 GTCCATCTTGGAGAGTCACTGTGGGAGGGCCAGAGTCCCAAGTGGGTTCAGTCCC 3671
 DB 364 laProGlyProAlaGlyProArgGlyAlaAlaGlyGlyGlyGlyGlyGlyGlyGly 379
 QY 3670 AGGCAGGGCTGGGGCAGGTAGCCAGAGCTGCCATGAGCCCGCCCAATCTGCCAG 3611
 DB 380 -----AspGlyValProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 396
 QY 3610 CAGGCCCTGGAGGCTGTGTGAGAAAGTGGGCCAGGCTGGGGCAGCTCAGCAGTGGCAGC 3551

QY 5059 GGGAACTGGGTGGACAGAGCTGTGTGGAGGACTGAGGGATCAAGGGTGTGGT 5000
 DB 69 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 84
 QY 4999 GGCAGCAGCAGTGGTAACTGCTGAGTGGCCCATCGCCCGAGTGGGGAGGAGGG 4940
 DB 85 GluSer-----GlyArgProGlyArgPro-----GlyAspArgGly 96
 QY 4939 ---CCTCTTCCCAGAGCTCTGTGCTGCTCCACAGAGGCAAGCAGGAGTG----- 4892
 DB 97 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 112
 QY 4891 ---GGAGCCAGGGTGGCAAGAGTCTGATGAGCAGCAGGAGGGGTGAGGCTGGGCAAGA 4835
 DB 113 ProGlyMetLysGlyHisArgGly---PheAspGlyArgAsnGlyGlyGlyGlyGlu--- 130
 QY 4834 GACTAAGTCTGGGGCAAGCATTTTATTGTTTAAATACAGAAATAGAAATTCGCAATAAA 4775
 DB 130 ----- 130
 QY 4774 TATCATCTAATAATAACATCTCCAAATAATAATAATAATAATAACACACTTAGAGTCA 4715
 DB 130 ----- 130
 QY 4714 TGAGTGGTGGGCTGGGGGCGAGGCGCTTGGGAGCTGCCA----- 4673
 DB 131 ---ThrGlyAlaProGlyLeuLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 149

Db	643	IleAlaGlyProArgGlySerProGlyGluArgGlyGluThrGlyProProGlyProAla	662
QY	2524	-----GCAGAAAGGAAGAAATGGAGCAGTCTCATGAGTAGGACAGAGAGAAAGG	2477
Db	663	GlyPheProGlyAlaProGlyGlnAsnGly	676
QY	2476	CTATGGAAAGACAGCATGTGTATAGGAAGAGGCAGCCCTTCCAAACCTCGGGGAGC	2417
Db	677	-----LysGlyGlu	679
QY	2416	GGCTGGGCCCCAGGAAAGGATGAGGGAGATGTATAACTGGGAAAGAGGAGAAAGCA	2357
Db	680	ArgGlyAlaProGlyGlu	688
QY	2356	GGGAGAGAGCCCTGGGATAGGCAAGGCCAGGGGAGGTTACACGCTGACAAAGCACAGA	2297
Db	689	GlyGlyProProGlyValAlaValProProGlyGlySerGlyProAlaGlyProProGly	708
QY	2296	GACAAACTAATTTCTCGGTGTATCTCGGGCTGGAGCCAGCTCCCCGCCACCTTCTTC	2237
Db	708	y-----ProGlnGlyValLysGlyLysArgGlySerProGlyGlyProGlyAl	724
QY	2236	CCTTTTCTTCTCCGCTCCAGAGTAGGCTCTGGGATGTGGAGAGAAACACGGGAGAG	2177
Db	724	alaGlyPhePro	731
QY	2176	GGGG---CCGGTTCCCATGTCCTCGCTCTCTCCAAACCTCTCTTCTCTCTCTCTGGG	2120
Db	731	GlyLeuProGlyProProGlySerAsnGlyAsnProGly---ProProGlyProSerGly	750
QY	2119	CTCTCTGAACACCTCTGGACACCCCCCAATCCTTACAGCATAACTCCATCCATCCACACCT	2060
Db	750	ySerProGlyLysAspGly---ProProGlyProAlaGlyAsnThrGly	765
QY	2059	AAACCCCTGCCACTACTCGCTGGGTTCCCTACAGTCCCTTCCCATGACCCCTCTAG	2000
Db	766	-----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaG	780
QY	1999	GAGGCCCATAGGTAGGAGGACAGCGAAGGAGGCGCTCTGGTCCAGCAGCAG	1940
Db	780	LysGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGly---AlaProG	798
QY	1939	GGCCC-----CAGGTCGGGGCAGAGCGAGCTGGAAACGCGCATGGGAGG	1892
Db	798	lyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetP	818
QY	1891	TGGGGCCCCCTC---TGTCCTCCACCCAAACCCCGCCGCGAAGACACCAACACCGTG	1835
Db	818	roGlyProArgGlySerProGlyProGlnGlyValLysGlyGlySerGlyLysProGlyA	838
QY	1834	TAAAC-----AGAGGAGACTTCCCGCCCCCGAGGGGAGCGGGC	1794
Db	838	laAsnGlyLeuSerGlyGluArgGlyPro---ProGlyProGlnGlyLeuProGlyLeuA	857
QY	1793	--GGCTGGGCTGGCAGCCAGTGGGCCAATGGGCCAAGCGGGTCCGAGGCTGCCACAT	1736
Db	857	laGlyThrAlaGlyLysProGlyArgAspGlyAsnProGlySerAspGlyLeuPro	875
QY	1735	GCCTAGTCTCTGGGCCACGGATGGGGAAGTGGGGGGCCACTAAGAGGGGAAGGCGAGA	1670
Db	876	-----GlyArg---AspGlySerProGlyGlyLysGlyAspA	887
QY	1675	GATAGACACAGAGGGGACTGTGTGGAGCGCAAGAGCCCTCTCTCCATCATTTTAC	1610
Db	887	rg-----GlyGluAsnGlySerProGlyAlaProGlyAlaProGlyHisProG	903
QY	1615	CAGAACCTCCACGGTTAGGACTGTCTTCCCCCAACTCTCTCTCTCTCTCTCTCTCT	1550
Db	903	lyProProGly-----ProValGlyProA	911
QY	1555	ATTCCTACGCTGGGATCTCATATAGTCCACAGCAAGCTGGCGAGTTGGTGTCTCCAGGTA	1499
Db	911	laGlyLysSerGlyAspArgGlyGlySerProGlyProAlaGlyProAlaGlyAlaProGlyP	931

QY 1495 GCGGTCAGGTATAGGTTTCTGTGATGAGGGG---CCAGGCCCTGGGTCCCTGTGCT 1440
 Db : : : : :
 QY 931 ro-----AlaGlySerArgGlyAlaProGlyProGlnGlyProArgg 945
 Db : : : : :
 QY 1439 GATTGAGAGTGGCACTGACGGAGGTCCAGAGCACCGGTGCACAGGACGCTAACATCC 1380
 Db : : : : :
 QY 945 lyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGlyPheP 965
 Db : : : : :
 QY 1379 CC-----CACGAGTCCCTGTGGCAGGATGACGAGAGCATG 1341
 Db : : : : :
 QY 965 roGlyAsnProGlyAlaProGlySerProGlyProAlaGlyGlnGlnGlyAlaIleGlys 985
 Db : : : : :
 QY 1340 AGCCGCGCCCG---GGCCCAACACGGGGAAGCAGCTGGCAGCGCCCTCAGGTGCTGTCCCC 1284
 Db : : : : :
 QY 985 exProGlyProAlaGlyProArgGlyProVal-----GlyProSerGlyProProGlyL 1003
 Db : : : : :
 QY 1283 ACCTTGGCTGGTGGGGGTGAGGT-----GGGACACAGATGGGAGGAGGAGGGA 1230
 Db : : : : :
 QY 1003 ysaPpGlyThrSer-GlyHisProGlyProIleGlyProGlyProArgGlyAsnArg 1022
 Db : : : : :
 QY 1229 TGGAGGAGGGGAAC-----ACTCCTGGCAAGACGGGAGAG 1194
 Db : : : : :
 QY 1023 GlyGluArgGlySerGluGlySerProGlyHisProGlyGln 1036
 Db : : : : :
 RESULT 12
 US-08-931-820-4
 ; Sequence 4, Application US/08931820
 ; Patent No. 6010863
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Assay for collagen degradation
 ; NUMBER OF SEQUENCES: 4
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/931,820
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP 96202596.1
 ; FILING DATE:
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1057 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; TISSUE TYPE: Collagen type III
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 1055
 ; OTHER INFORMATION: /label= Modified
 ; OTHER INFORMATION: /note= "Ala may be Pro"
 US-08-931-820-4

Alignment Scores:
 Pred. No.: 1,26e-15 Length: 1057
 Score: 327.50 Matches: 326
 Percent Similarity: 29.96% Conservative: 80
 Best Local Similarity: 24.06% Mismatches: 501
 Query Match: 3.51% Indels: 450
 DB: 3 Gaps: 71

US-09-931-704-3 (1-5087) x US-08-931-820-4 (1-1057)

QY 5059 GGGAAACCTGGGTGGACAGAACTGTGTGGAGGACTGAGGGATCAAGGGTTGGTGT 5000
 Db : : : : :
 QY 68 GlyProProGly-----AlaIleGlyProSerGlyProAlaGlyLysAspGly 83
 Db : : : : :
 QY 4999 GGACGACGAGTGGTAACTGTGAGGTGCGCCATCCCGCAGTGCAGGAGGAGGAGG 4940
 Db : : : : :
 QY 84 GluSer-----GlyArgProGlyArgPro-----GlyGluArgGly 95
 Db : : : : :
 QY 4939 ---CCTCTTCCAGAGCTGTGTGAGCTCCACAGAGGCAAGCAGGAGTGT----- 4892
 Db : : : : :
 QY 96 LeuProGlyProProGlyIle-----LysGlyProAlaGlyIleProGlyPhe 111
 Db : : : : :
 QY 4891 ---GGAGCCAGGGGTGCGCAAGACTCTGATGAGCAGGAGGAGGTGAGGCTGGSCAAAGA 4835
 Db : : : : :
 QY 112 ProGlyMetLysGlyHisArgGly-----PheAspGlyArgAsnGlyGluLysGlyGlu 129
 Db : : : : :
 QY 4834 GACTAAGTTCTGGGGCAAGCATTTTATTGTTTAAATACAGAAATAGAAATCTGCAATAAA 4775
 Db : : : : :
 QY 129 ----- 129
 QY 4774 TATCATCTAATAATAACATCTCCAAATAAATAATTAATTAACAACACACTTAGAGTCA 4715
 Db : : : : :
 QY 129 ----- 129
 QY 4714 TCAGTGGGTGGGCTGGGGGCGAGGCGCTTGGGAGCTGCCA----- 4673
 Db : : : : :
 QY 130 ---ThrGlyAlaProGlyLeuLysGlyGluAsnGlyLeuProGlyGluAsnGlyAlaPro 148
 Db : : : : :
 QY 4672 ---CCCTACACCCCAAAATGCTACTGCAATTGTAAACTTTTCAGGAAATCTGTGTGTGGC 4616
 Db : : : : :
 QY 149 GlyProMetGlyProArg----- 154
 Db : : : : :
 QY 4615 TATGTTGGCTTCCCGCAGGCTGGCAACCCACAGATACCTTGGGAAAGGGGCGAGAGAAGA 4556
 Db : : : : :
 QY 155 ---GlyAlaProGlyGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArg--- 172
 Db : : : : :
 QY 4555 GGCACCGTAATACTGGAGCAAAATTCGAGGCAAGGTCTGTGATGCTCAGCTCGGTTTTGT 4496
 Db : : : : :
 QY 173 -----GlyAsnAspGlyAlaArgGlySerAspGlyGlnProGlyProPro 187
 Db : : : : :
 QY 4495 TTGCCACTGTGTGTTTGGAGACTCTGGGTCTGGTCTGTCTGTCTGCCCTGCCCTGTGCT 4436
 Db : : : : :
 QY 188 GlyProProGlyThrAlaGlyPheProGly----- 197
 Db : : : : :
 QY 4435 TTCAGTATTTCTCTTCTGTTTATAGGCGACCTCCAATGCACACCCCAACCTGTGACTG 4376
 Db : : : : :
 QY 198 -----SerProGlyAlaLysGlyGluValGlyProAlaGlySerPro----- 211
 Db : : : : :
 QY 4375 TTAATCTCATCTCTTTTGTGTAGAATTGCCACCATGTTTGTCTGTAATGATGTAT 4316
 Db : : : : :
 QY 211 ----- 211
 QY 4315 AGAGTGATTGGGATAGGGAGCAACCCATCCAAGAGAGTGGGCGAGCCAAAGGCG 4256
 Db : : : : :
 QY 212 -----GlySerAsnGlyAlaProGlyGlnArgGlyGluPro----- 223
 Db : : : : :
 QY 4255 TTCACCCCAAGAGCATGTGTGGTGGCGCCCTCAACCCCTGCCCTCCAGATGTGCCA 4196
 Db : : : : :
 QY 224 ---GlyProGlnGlyHisAlaGly-----AlaGlnGlyProProGly----- 236
 Db : : : : :
 QY 4195 CCTGAGGGGCTGGGTGGGCACTGGCCAAAGTGTAGCAAGAGAGGAGGAGGAGGAGGAG 4136
 Db : : : : :
 QY 237 ProProGlyIleAsnGlySer---ProGlyGlyLysGlyGluMetGly----- 251
 Db : : : : :
 QY 4135 CTCTGCACCAACCTGAACCACTTCACATCTCCCTCGAGCATGACTTCTGTAGAAACAGGA 4076
 Db : : : : :
 QY 252 -----ProAlaGlyIleProGlyAla-----ProGlyLeuMetGly 263
 Db : : : : :
 QY 4075 CAGGACAGGGCTTACTGTACTCTCTCCCGAGCTCGGTAGACCT-----TTGGGAGGTGG 4022
 Db : : : : :
 QY 264 -AlaArgGlyProProGlyProAlaGlyAlaAsnGlyAlaProGlyLeuArgGlyGlyAl 283
 Db : : : : :
 QY 4021 GGAGGAGACAGGGGTGATCGCATCAACAGCCCGCGGTCCAGGAAAGGGCCAGAGGCTC 3962
 Db : : : : :
 QY 4021 GGAGGAGACAGGGGTGATCGCATCAACAGCCCGCGGTCCAGGAAAGGGCCAGAGGCTC 3962

Db 283 aGlyProGlyLysAsnGlyAlaIalysGlyGluProGlyProArgGlyGluArgGly-- 302
Qy 3961 ACAGCTTCTGCTCCCTGGCTCAACAGGTGTT---GCCATACAGGGTGGCTCTCAACAAG 3905
Db 303 -----GluAlaGlyIleProGlyValProGlyAlaIalysGly-GluAspGlyLysA 319
Qy 3904 TGGAGCAGGGTTTGAAGGGGAGCGAAGAGAGAGAGGTTCAGAGCTCAGNAGCCATCAGC 3845
Db 319 spGlySerProGlyGluProGlyAlaAsnGlyLeuProGlyAlaIalysGlyGluArgGlyA 339
Qy 3844 CCCAGGT-----GCAGGTGAGTGCAGCTGCTGGAGGTGCATCTTCTTCTT 3797
Db 339 laProGlyPheArgGlyProAlaGly----- 347
Qy 3796 GAGCCGGTTGAAGTCTTGGCCGAGCGCCACAGCCAGGTCTGAGCTCTTTCAGAGCCCA 3737
Db 348 -----ProAsnGlyIleProGlyGluLysGlyProAlaGlyGluA 361
Qy 3736 GAAGTCGTCATCTTGGAGGAAGTCACTGTGGGCGAGGCCAGGAGTCCAGTGGGTTTC 3677
Db 361 rgGlyAlaProGlyProAlaGlyProArgGlyAlaIalysGlyGluProGlyArg----- 378
Qy 3676 AGTCCAGGCGCGCTGGGCGAGTGGTAGCCAGAGTGCATCAGCGCCGCAATGCT 3617
Db 379 -----AspGlyValProGlyGlyProGlyMetArgGlyMetProGly---S 393
Qy 3616 GCCCAGAGCCCTGAGGCTGTGTGAGAGTGGCCAGAGTGGCGCGCAGCTCAGCAGT 3557
Db 393 erProGlyGlyProGlySerAspGlyLysProGlyProArgGlySerGlnGlyGluSerG 413
Qy 3556 GCGAGCTGACGTTGAGGCCAGCGCAAGTAAACAGAGAGTGGCTGAGGCTCGTAGTT 3497
Db 413 lyArgPro-----GlyProProGlyProSerGlyProArg----- 424
Qy 3496 CTGGGTGAGCGCGAGTTTGTATTGAGGCTTGGCCACACCTCCCAAGTCAACAGTGGCCCT 3437
Db 425 --GlyGlnProGlyValMet-----GlyPheProGlyPro-Lys----- 436
Qy 3436 GGGCAGAGTCTGCCCCCAGGGGAGGAGTGAAGTCTGGCTGTTGAAGGGGGGCC 3377
Db 437 -GlyAsnAspGlyAlaProGlyLysAsnGlyGluArgGlyGly-----ProGlyGlyPr 454
Qy 3376 CAGGTAGTTGAGCTGTGAAAAGGAGGAGGTGATGCGGAAGGAGGAGGCGAGCGCTGG 3317
Db 454 o-----GlyProGlnGlyProProGlyLysAsnGlyGluThrGlyProG 469
Qy 3316 CTCACCACCAAGGAGATACCTGCTCCCAAGTTTCTATTTTGTGTCCT----- 3265
Db 469 nGlyProPro-----GlyProThrGlyPr 477
Qy 3264 -----CACACTGCCCCGTCTCCATGCTAGGCTCTTGTTCATGGCT 3221
Db 477 oGlyLysAspLysGlyAspThrGlyProProGlyProGlnGly-----LeuGlnGlyLe 495
Qy 3220 CCCTGAGTTTTTCCAATAAGGATATCATTTGTATGCAGGTGACAGTTGAAGCAGTGGCGG 3161
Db 495 uPro-----GlyThrGlyG 500
Qy 3160 TGGTGAGCTCTCCAATGAGAGGCGTA-GAAAAAGGAGGAGGAGGAGGAGCGCACAGAGCGC 3102
Db 500 yProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAla----- 517
Qy 3101 TCAGGAGGAGCTGCTGATTGGCTGGGTGGGCAACCTTGGGAGAGTCCCCGTGTGGG 3042
Db 517 ----- 517
Qy 3041 GCAGGACAGAGGAGGAGCTGGAGAGGAGTGGAGAGGTAGAGTGGCCCGCAGAGA 2982
Db 518 -----GlyAlaProGlyAlaProGlyGlyLysGlyAspAla---GlyAlaProGlyG 534
Qy 2981 GTGAGAGACCAAGAGAGGAGTTCGAGAGGAGGAGCGCTCAGCAGTGTGGCTGCNA 2922

534 uArgGlyProGlyLeuAlaGlyAlaProGly----- 545
2921 AACAGACAGCAAGGAGAGTACAGCTTCAAAACACTGTGCGGCTTGGCAAGGAGGTGCCTG 2862
546 -----LeuArgGlyGlyAlaG 551
2861 ACACCTTGGAGAGAACCATTTCTGGGCGCTTGGAAAGCAGAGTTCAGGGGGTTAAGAT 2802
551 yProPro-----GlyPro-----GluGlyGlyLysGlyAl 561
2801 GAGTGGTGGAGAGAAAGTGGAGCAGAGGGGTGTAGACCACTCTTGGAGAAATTTGGCA 2742
561 aAlaGlyProGlyProGlyAlaAlaGlyThrProGlyLeu-----G 577
2741 GGGAAAGCCAGACAGAAA-----AAGGTTCGGAACTAGA 2706
577 nGlyMetProGlyGluArgGlyGlyLeuGlySerProGlyProLysGlyAspLysGlyG 597
2705 GAGGCGAGTGGAGTGAACAAAGTTTTCACAAAGAAACAGAGAGAGATGAGAGTGG 2646
597 uProGlyGlyProGlyAlaAspGlyValProGlyLysAspGlyProArgGlyProThrG 617
2645 CCCATTGGAACACAGGCCCCATTTCCTGCTGACACACCCAGGCCACAGCTGAGGGGCT 2586
617 yProIleGlyProProGly-----ProAlaGlyGlnProGlyAspLysGlyGlyG 634
2585 GGGGAGAGACGGGTACTCACTAAACACTCCGAAAGAGGGTTTCAGCTGATAGGACAGGCT 2526
635 -GlyAlaProGlyLeuProGlyIleAlaGlyProArgGlySerProGly---GluArgG 653
2525 TGCAGAGGAAGAAATGAGGAGCTCTCATGATGAGGACAGGAGGAGGAGCTATGGAAGAC 2466
653 yGluThrGlyProProGlyProAlaGlyPheProGlyAlaProGlyGln-AsnGlyGluP 673
2465 AGCATGTTATAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2406
673 roGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 685
2405 AGGAAAGAGGAGTGGAGAGATGTATACTTGGGAAAGAGGAGGAGGAGGAGGAGGAGGAG 2346
685 yGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 700
2345 TGGATAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2286
700 erGlyProAlaGlyProProGly----- 707
2285 TTCCTCGGTGATCTCTGGGCTGGAGCCAGCTCCCGCCAGCTCTTCTTCTTCTTCTTCTT 2226
708 --ProGlnGlyValLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 727
2225 CCTGCTCCAGAGTAGGCTCTGGGATGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2169
727 ro-----GlyAlaArgGlyLeuProG 734
2168 TTCCTCATGTCCTCGGCTCTCCACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2109
734 yProProGlySerAsnGlyAsnProGly---ProProGlyProSerGlySerProGlyL 753
2108 CCTCTGGAGCCACCCCAATCTCTTACAGCATACCTTCCATCCACACCTTAAACCCCTGC 2049
753 yAspGly---ProProGlyProAlaGlyAsnThrGly----- 764
2048 CACTACTGCTGGTTCCTTACAGTCCCTTCCCATGACCCCTTACAGGAGGAGGAGGAGG 1989
765 ----AlaProGlySerProGlyVal-SerGlyProLysGlyAspAlaGlyGlnProGly 782
1988 AGGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1935
783 GluLysGlySerProGlyAlaGlnGlyProProGly-----AlaProGlyProLeuGly 800
1934 -----CAGTCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1881
801 IleAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetProGlyProArg 820

Db	60	GlyGluGlnGlyProArgGlyAspArgGlyAspLysGlyGluLysGlyAlaProGlyPro	79
Qy	4245	AAAGCATGTGGT-----GGTGGCCCTCAACCCCTGC-CCTCCAGAT	4202
Db	80	ArgGlyArgAspGlyGluProGlyThrLeuGlyAsnProGlyProProGlyPro	99
Qy	4201	GTGCCA---CCTGAGGGGCTGGTGGCACTGCCCAAGTGGTAGGCAAGACGAGGCGC	4145
Db	100	ProGlyProProGlyLeuGlyGlyAsnPheAlaAala-----GlnMetalaGlyGly	117
Qy	4144	CATGAGCGCTCTGC-----ACCAACCTGAACCACTTCACACTCCCTCGAGCATGACTT	4091
Db	118	PheAspGluLysAlaGlyGlyAlaGlnLeuGlyValMetGlnGlyProMetGlyPro	137
Qy	4091	-----	4091
Db	138	GlyProArgGlyProProGlyProAlaGlyAlaProGlyProGlnGlyPheGlnGlyAsn	157
Qy	4090	-----CCTGTAGAAACAGGACAGGACAGGCGCTACTGTACTCTCTCTCC	4049
Db	158	ProGlyGluProGlyGluProGlyValSerGlyPro-MetGlyProArgGlyProProGly	177
Qy	4048	CCAGCTCGGTAGACCTTTGGGAGGTGGAG-----GAGACAGGGCTGATCGCATCACA	3995
Db	177	yProGlyLysProGlyAspAspGlyGluAlaGlyLysProGlyLysAlaGlyGluAr	197
Qy	3994	CGCCACGCGGTCCAGGAAGGCGCAGAGCTCACAGCTTCTGCTCTCGCTCGCTCAACAGG	3935
Db	197	gGlyProProGlyProGlnGlyAlaArgGly-----PheProGlyThrProGly	213
Qy	3934	TGTT---GGCATACAGGCTGCTCTCACAAAGTGGAG-CAGGGTTTGAAGGGGAGCG	3879
Db	213	yLeuProGlyValLysGlyHisargGlyTyProGlyLeuAspGlyAlaLysGlyGluAl	233
Qy	3878	AAGAGAGAAGTCAAGACTCAGAACCATGACCCCGAGTGCAGGTCAGCTGCAGCTG	3819
Db	233	aGlyAlaProGlyValLysGlyGluSerGlySer--ProGly-----	246
Qy	3818	CTGGAGGTCATCTTCTTTGAGCGGGTTGAAGTCTTGGCGGAGCGCCACAGCCAGG	3759
Db	247	---GluAsnGlySerProGlyProMetGlyProArgGlyLeuProGlyGluArgGlyArg	265
Qy	3758	TCTGCAGCTCTTCAGCAGCAGAAAGTCTGTCCATCTCTTGAGGAAAGTCACTGTGGG	3699
Db	266	ThrGlyProAlaGlyAlaAlaGlyAlaArgGlyAsnAspGly-----	279
Qy	3698	GGCCAGGATCAAGTGGGTTTCACTCCAGGACGGCTGGGAGTGGGTAGCCACAGAG	3639
Db	280	-----GlnProGlyProAlaGlyProProGly	288
Qy	3638	CTGCCATGACGCCGCAATGCTGCCAGCAGGCGCTGGAGCTGGTGCAGAAGTGGGCCA	3579
Db	289	-----ProValGlyProAlaGly-----GlyPro	296
Qy	3578	GGCTGGCGGACACTCAGCAGTGGCAGCTTGAACCGTTCAGGCCACGCAAGTAAACAGAA	3519
Db	297	GlyPheProGlyAla-----ProGlyAlaLysGlyGluAlaGlyProThr---	311
Qy	3518	GGTGGCTGTAGGCCTCGTAGTTCTGGGTACGCGCAGTTTGTCAATGAGGCTGCCACA	3459
Db	312	-----GlyAlaArgGly	315
Qy	3458	CCTCCAAGTCAACAGTGGCCCTGGGACAGTCTCTGCCCCAGCGGGAGGTTGAAGT	3399
Db	316	ProGluGlyAlaGlnGlyProArgGlyGluProGlyThrProGlySerProGlyProAla	335
Qy	3398	CTGGCTCGTTGAAGGGGGGCCAGCTAGTGTAGCTGTGTGAAGGAGAGGCTGATGGGGA	3339
Db	336	GlyAla-----SerGlyAsnProGlyThrAspGlyIlePro-----GlyAlaLysGly	351
Qy	3338	AGGAGGAGGCGAGCGCTGGTCTCACCCAAAGGAGATACCTGCTCCCAAAGTTTCTAT	3279
Db	352	SerAtaGlyAlaProGlyIleAlaGlyAlaProGly--PheProGlyProAr-----	368

Qy	3278	TTTTTGTC	CCCCTGACACTGGCCCTGTCTCCATGCTAGGCTTCTTTGTTCATGGCGCTCC	3219
Db	369	-----	-----GlyProPr	371
Qy	3218	CTGAGTTTTTCC	ATAAGGATATCATTTGTATGTCAGGTGACAGTTGAACGAGTGGCGGTG	3159
Db	371	o-----	-----	371
Qy	3158	GTGAGCTCTCCA	ATGAGAGGCGTAGAAAAAGGAGGAGGAGGAGCGCACAAAGACGCTCA	3099
Db	372	-----	-----AspProgl	374
Qy	3098	GGAAGCAGCTGCTG	ATTGGCTGGT-----GGGCAAAACC-----TTGGGAGAGTGC	3051
Db	374	n---	GlyAlaThrGlyProLeuGlyProLysGlyGlnThrGlyLysProGlyIleAlaGl	393
Qy	3050	GCTGTGGGGCAGG	CAGAGAGGAAAGAGGAGCTCGAGAGCTGGAGAGTAGTAAGGTGG	2991
Db	393	yPheLysGlyGluGlnGly	ProLysGlyGluProGlyProAlaGlyPro---GlnGlyAl	412
Qy	2990	CCCAGAGTGCAG	AGACCAAGGA---AGAGGAGTTCCAGAAAGCGCAGCAGCGCTCAGCA	2934
Db	412	aProGlyProAlaGly	GluGluGlyLysArgGlyAlaArg-----	425
Qy	2933	GTGTGGCTGCAAA	ACAGACAGCAAGAGGAATCAGCGCTTCAAAACACTGTCTGGGCTTGGCA	2874
Db	426	-----	-----GlyGluPro-----	428
Qy	2873	AGGAGTGCTGAC	CGACCTTGGAGAGAACCATGTCTGGGGCTTGGAGCCAGATTGCAGG	2814
Db	429	-GlyGlyVal-----	-----GlyProIleGlyProProGlyGluArgGlyAlaPr	443
Qy	2813	GGCTTAAGAAGT	AGCTGGGTGAGAGGAAGTGGAGCAGCGGTGTAGACCACTCTCTGG	2754
Db	443	oGly-----	AsnArgGlyPheProGlyGlnAspGly-----LeuAlaGl	456
Qy	2753	AGAAATTTGGCAG	GGAACCCAGGAGAGAGAAAAAGGTTCGGAACTAGAGAGGCGAGTGA	2694
Db	456	yPro-----	-----LysGlyAlaProGly-----	462
Qy	2693	GTGAAGCAAAAGG	TTTTTCTGCACAAGGAACAAGAGAGAGATGAGAGTGGCCCATTTGGAACC	2634
Db	463	-----	-----GluArgGlyProSerGlyLeuAlaGlyPr	472
Qy	2633	A-----	GAGCCCATTTGCCGTGTGACACACCAGGC---CCACCAGCTGA	2592
Db	472	oLysGlyAlaAsnGly	AspProGlyArgPro---GlyGluProGlyLeuProGlyAlaAr	491
Qy	2591	GGGGCTG---	GGGAGAGACGGGTACTCATACTCCGAAAGAGGGTTCAGCTGATAG	2535
Db	491	gGlyLeuThrGlyA	rGProGly-----	498
Qy	2534	GAGCAGGGTTGC	AGAAGAAATGAGGCGACTCTCATGAGTAGGACAGGAGGAAGGCT	2475
Db	499	-AspAlaGlyProGlnGly	LysValGlyProSerGlyAlaProGlyGluaspGlyArg-p	518
Qy	2474	ATGGAAGACAGC	ATGTTATAGGAAGAGGCGCAGGCC-----TTTCCAAACC	2427
Db	518	roGlyProProGlyProGlnGly	AlaArgGlyGlnProGlyValMetGlyPheProGlyP	538
Qy	2426	TCGGGGGAGCGCT	GGGGCCAGGGAAGGATGAGG-----	2389
Db	538	roLysGlyAlaAsnGly	GluPrroGlyLysAla---GlyGluLysGlyLeuProGlyAlap	557
Qy	2388	-----	AGATGTATACTGGAAAGAGGAGAACGAGGCGAGGC-----	2347
Db	557	roGlyLeuArgGlyLeuProGlyLys	AspGlyGluThrGlyAlaGluGlyProProGlyP	577
Qy	2346	--	CTGGATAGGCAAGGGCCAGGGGAGGTTTCACACGCTGACAGCACAGACAAACT	2289
Db	577	roAlaGlyProAlaGly	GluArgGlyGlu-----GlnGlyAlap	590

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:49:37 ; Search time 24.1925 Seconds
(without alignments)
6508.980 Million cell updates/sec

Title: US-09-931-704-4
Perfect score: 1498
Sequence: 1 tattattaaagcttcgcgg.....aggccacagtcagctgtgctt 819

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO spo1/US09931704/runat_27012003_154126_3648/app_query.fasta_1.7189
-DB=PIR_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704 @CGN_1_1_351 @runat_27012003_154126_3648 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	145	9.7	325	2 T32248	hypothetical prote
2	142.5	9.7	660	1 Q0B3	BHLF1 protein - hu
3	141	9.4	383	2 S32975	gene BCRF2 protein
4	137	9.1	319	2 T32250	hypothetical prote
5	137	9.1	574	2 T43556	Wiskott-Aldrich sy
6	137	9.1	574	2 T38819	wiskott-aldrich sy
7	136.5	9.1	1146	2 A38587	collagen, cornea s
8	135	9.2	371	2 E88633	protein F56B3.1 [i
9	135	9.0	1691	1 S32917	collagen alpha 5(i
10	135	9.0	1763	2 S16366	collagen alpha 2(i
11	134	8.9	302	2 T15936	hypothetical prote
12	134	9.2	627	2 A44112	spidroin 2, dragli
13	133	8.9	304	2 T2482	hypothetical prote
14	132	8.8	1433	2 A46053	bullous pemphigoid

15	132	8.8	1669	1 CGMS4B	collagen alpha 1(I
16	131.5	8.8	303	2 T28999	hypothetical prote
17	131	8.7	317	2 T19143	hypothetical prote
18	131	8.7	754	2 A5267	collagen alpha 5(I
19	130.5	8.7	517	2 T10927	3C3.18c protein -
20	130.5	8.7	677	2 S23296	collagen alpha 2(I
21	130	8.7	660	1 Q0B3	BHLF1 protein - hu
22	130	8.7	1532	2 A61262	collagen alpha 1(X
23	129.5	8.6	304	2 T22602	hypothetical prote
24	129.5	8.6	380	2 T28888	cuticle collagen d
25	129.5	8.6	1670	1 CGHU3B	collagen alpha 3(I
26	129	8.6	316	2 T12888	hypothetical prote
27	129	8.6	576	2 T36729	probable serine/th
28	129	8.6	1147	2 T42627	ADP-ribosylation f
29	128.5	8.6	304	2 T26185	hypothetical prote
30	128.5	8.6	360	2 T26184	hypothetical prote
31	128.5	8.6	360	2 T37285	collagen dpy-2 - C
32	128.5	8.6	688	2 A53330	collagen alpha 2(I
33	128.5	8.6	1541	2 T02831	AAA protein L4171
34	128	8.5	299	2 T19564	hypothetical prote
35	128	8.5	316	2 T19291	hypothetical prote
36	128	8.5	551	2 S57447	HP3811-7 protein -
37	127.5	8.7	316	2 S08169	collagen col-12 pr
38	127.5	8.7	316	2 S08170	collagen col-13 pr
39	127.5	8.5	1250	1 B45219	N-methyl-D-asparta
40	127	8.7	255	2 J00320	hypothetical 24.7K
41	127	8.7	291	2 T26576	hypothetical prote
42	127	8.5	1255	2 T31065	diaphanous protein
43	126.5	8.4	299	2 T29956	hypothetical prote
44	126	8.4	615	2 A05269	collagen alpha 1(I
45	126	8.4	1008	2 T04462	hypothetical prote

ALIGNMENTS

RESULT 1

T32248
hypothetical protein T15B7.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000

C/Accession: T32248

R/Pauley, A.; Gattung, S.

submitted to the EMBL Data Library, September 1997

A/Description: The sequence of C. elegans cosmid T15B7.

A/Reference number: Z21139

A/Accession: T32248

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 1-325 <PAU>

A/Cross-references: EMBL:AF022985; PIDN:AAB69961.1; GSPDB:GN00023; CBSP:T15B7.4

A/Experimental source: strain Bristol N2; clone T15B7

C/Genetics:

A/Gene: CBSP:T15B7.4

A/Map position: 5

A/Introns: 266/1

C/Superfamily: unassigned collagens

Alignment Scores:

Pred. No.:	0.000832	Length:	325
Score:	145.00	Matches:	73
Percent Similarity:	32.94%	Conservative:	11
Best Local Similarity:	28.63%	Mismatches:	94
Query Match:	9.68%	Indels:	77
DB:	2	Gaps:	11

US-09-931-704-4 (1-819) x T32248 (1-325)

QY 4 TATTAAAGCTTCGCGGAGCGCGCTCGCCTCCACTCGCCAGCTCTGGG----- 57

Db 49 TyrGluGlnIleAlaTrpGlnAlaMetIleProThrArgProSerSerGlySerSer 68

QY 58 -----AGAGGA 63

|||||

Db	204	SerProSerArgArgGlyAla-----SerLeuGlyProGlnValGlnProHis	219
QY	385	CAATGACAGCTGCGCTGACCCAGAACTATAGGCGGTACAGTCACCTCTCTGTGTTACTT	444
Db	220	ArgAspProSerGlyProAspProProThrGlyProSerLeuGlyProProAlaProLeu	239
QY	445	CGGTGGCTCAACCGTCAGGCTCCACAGCTGACCTCCGACGTAGCTGGCCCACTTCGT	504
Db	240	-----GlnProSer---LeuHis-----ProArgProGlnLeu	249
QY	505	TACAGCCTCCAGGGCGCTGTGGCGAGCATTTGACGGTGTATGGCGACGCTTGCGTACCC	564
Db	250	LeuAlaSerProGlyProProGlyGln-----pro	259
QY	565	ATGCCCCAGCCTCTGCCAGGAGTACGCCAGCCTGGGCCCTGG-----	609
Db	260	GluGlyProArgGlnProGlyArgValAlaPheProLeuProTrpProLeuProAla	279
QY	610	-----CCCTGCCACAGTACTTCTCTCCA-----GAA	636
Db	280	SerHisProSerProLeuSerLeuProHisArgValHisGlnAlaGlyArgArgAsp	299
QY	637	GATGATGACTTCTGGCTGCTGAAGAGAGCTGCAGACCTGGCTATGGCGTTTCAGCCAGGA	696
Db	300	ProGly-----GlyProValSerValProProAlaAlaAlaGlnSer	313
QY	697	CTTCAACCGGCTTAAGAAGAGATGCAGCGCTTCAGTCAACCTGCACCTTGGGA	756

```

Db      314 LeuProPro-----GlyLysGlyAlaSerPheSer----- 323
QY      757 GGCACATGGTTTCTGACCTCTGACCCCTTAACCCCCACACCTCCAGGCCCCAGTCAGCTGTG 816
Db      324 ||||| ||| ||||| |||||
          ProProSerLeu-ArgProSerLeuLeuCy 333

QY      817 C 817
Db      333 s 333

RESULT 4
T32250
hypochothetical protein T15B7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32250
R:Pauley, A.; Gattung, S.
submitted to the EMBL data library, September 1997
A:Description: The sequence of C. elegans cosmid T15B7.
A:Reference number: Z21139
A:Accession: T32250
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-319 <PAU>
A:Cross-references: EMBL:AF022985; PIDN:AB69959.1; GSPDB:GN00023; CESP:T15B7
A:Experimental source: strain Bristol N2; clone T15B7
C:Genetics:
A:Gene: CESP:T15B7.3
A:Map position: 5
A:Introns: 266/1
C:Superfamily: unassigned collagens

Alignment Scores:
Pred. No.: 0.00367 Length: 319
Score: 137.00 Matches: 69
Percent Similarity: 31.30% Conservative: 13
Best Local Similarity: 26.34% Mismatches: 95
Query Match: 9.15% Indels: 85
DB: 2 Gaps: 10

US-09-931-704-4 (1-819) x T32250 (1-319)

QY      4 TATTAAAGCTTCGGCGGAGCGCGCTCCCACTCCGCGCCAGCCTCTGGG----- 57
Db      49 TyrGluGlnIleAlaTrpGlnAlaMetIleProThrThrArgProSerGlySerSer 68
          ||||| ||||| ||||| ||||| ||||| |||||

```


					:::	
Db	469	uProProAlaAalProAlaProAlaProProAlaProAlaProAlaProValAl	489			
Qy	625	-----CTTCCTCCAGAAATGGATGACTT-----	CTGGCTGC	656		
Db	489	aserIleAlaGlLeuProGlnAspGlyArgAlaAsnLeuMetAlaSerIleArgAl	509		:::	
Qy	657	TGAAGGAGCTGCACAGCTGCTATGGGTTCAGCCCAAGAGCTTCAACC	GGCTTAGAGA	716	:::	
Db	509	aserGlyGlyMetAspLeuLeuLysSerArgLysValSerAlaSerProSer	-----	526		
Qy	717	AGATGCAGCCTCCAGCAGCTTCAGTCACCCCT	747			
Db	527	----ValAlaSerThrLysThrSerAsnPro	535		:::	

A03587
collagen, cornea-specific - chicken
C/Species: Gallus gallus (chicken)
C/Date: 14-Feb-1992 #sequence_revision 15-Aug-1997 #text_change 20-Sep-1999
C/Accession: S16501; A38587
R/Marchant, J.K.; Linsenmayer, T.F.; Gordon, M.K.
Proc. Natl. Acad. Sci. U.S.A. 88, 1560-1564, 1991
A/Title: CDNA analysis predicts a cornea-specific collagen.
A/Reference number: A38587; UID:91142213; PMID:1705041
A/Accession: S16501
A/Molecule type: mRNA
A/Residues: 1-1146 <MAR>
A/Cross-references: EMBL:M60172; NID:G211609; PID:AAA48703.1; PID:G211610
A/Accession: A38587
A/Molecule type: mRNA
A/Residues: 1-174, 'X', 176-233, 'X', 235-344, 'X', 346-408, 'X', 410-499, 'X', 501-
A/Cross-references: GB:M60172
C/Superfamily: unassigned collagens
C/Keywords: Cornea

Alignment Scores:			
Pred. No.:	0.00393	Length:	1146
Score:	136.50	Matches:	80
Percent Similarity:	34.35%	Conservative:	21
Best Local Similarity:	27.21%	Mismatches:	102
Query Match:	9.11%	Indels:	91
DB:	2	Gaps:	17
US-09-931-704-4 (1-819) x A38587 (1-1146)			
QY	34	CTCTCCCACTCCGCAGCCTCTGGAGAGAGCCGCGCC	---GGCGCGCCCGCCCCCAG 90
Db	240	ProProGlyProProGlySerAlaGlyLeuLysGlyProMetGlySerProGlyProGln	259
QY	91	---CCCCATGCAGCTCCGAGC---AGCGGACTCGTGGGGGATGTTAGCTTGCT	-----138
Db	260	GlyProGlyProProGlyLeuGlnGlyPheArgGlyGluAlaGlyLeuProGlyAla	279
QY	139	-----ATGCAGGTGCTGTGGCACCCTCCCTGCAGTGCAGCTCTTAATCGCACAGAGA	192
Db	280	LysGlyGluLysGlyAlaThrGlyProProGlyProLysGlyAspGlnGlyGluLysGly	299
QY	193	TCAGGCCCTGGCCCTCCATCCAGAAAACCTATGACCTACCCGCTACCTGGAGATCA	252
Db	300	AlaHis-----GlyMetThr	304
QY	253	ACTCGCGAGCTTAGCTGGGACCTACCTGAACCTACCTGGGGCCCTTTTCAACGAGCCTGA	312
Db	305	GlyGluGlnGlySerArgGlyIleProGlyProProGlyGluProGlyAlaLysGly	323
QY	313	CTTCAATCTCTCCGCTGCGGGGCGCAAAACTCTGCCAGGCG	---CACGGTCAACTTGA 369
Db	324	-----ProAlaGlyGlnAlaGlyArgAspGlyGlnProGlyGluArgGlyGluProGly	341
QY	370	AGTGTGGCGAAGCCTCAATGACAGGCT	-----GCGGCTGACCCA 408
Db	342	LeuMetGlyIleProGlyAlaArgGlyProProGlyProSerGlyAspThrGlyGluPro	361


```
QY 409 GAACTATGAGCGGTACAGTCACCTCTCTGTGTTACTTTCGCTGGCTCAACCGTCAGGCTGC 468
Db 362 GlyLeuThrGlyProGlnGlyProProGlyLeuProGlyAsnProGlyArgProGlyAla 381
QY 469 CAC-----AGCTGAACCTCCGACG 486
Db 382 LysGlyGluProGlyAlaProGlyLysValIleSerAlaGluGlySerThrIleAla 401
QY 487 TAGCTGCGCCCACTTCTGTACAGCTCCAGGCGCTGCTGGCAGCATTTGCAGGTGTCAT 546
Db 402 LeuProGlyPro-----ProGlyProProGlyProIleGly----- 413
QY 547 GCGCAGCTTGCTACCTCCCACTGCCAGCT---CTGCAGGAGACTGAGCCAGCTGGGC 603
Db 414 -----ProThrGlyPro-ProGlyValProGly-----ProVal-GlyP 426
QY 604 CCTGCGCCCTGCCCACTGACTTCTCCAGAAGATGATGACTTCTGCTGCTGAAGGA 663
Db 426 roAlaGlyLeuPro----- 430
QY 664 GCTGACAGCTGGCTATGCGGTTTCAG---CCAAGGACTTCAACCGGCTTAAGAGAAGAT 720
Db 431 -----GlyGlnGlnGlyProArgGlyGluLysGlySerAlaValGluV 445
QY 721 GCAGCTCCAGCAGCTTCACTCACCTCGCACT-----TGGAGGCACATGTTTCTTGACC 774
Db 445 alValIleGluThrIleYsThrGluValSerSerLeuAlaSerGlnMetLeuSerAspL 465
QY 775 TCTGACCTTTAACC-----CCACACCTCCAGGCCCA 806
Db 465 euGlnGlyArgAlaGlyProProGlyProProGlyPro 477
RESULT 8
E88633
protein F56B3.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: E88633
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88633
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-371 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAO2612.1; PID:g2854198; GSPDB:GN00022; CESP:F56B3.1
A;Note: contains similarity to collagens
C;Genetics:
A;Map position: 4
C;Superfamily: unassigned collagens
Alignment Scores:
Pred. No.: 0.00531 Length: 371
Score: 135.00 Matches: 71
Percent Similarity: 35.68% Conservative: 10
Best Local Similarity: 31.28% Mismatches: 99
Query Match: 9.23% Indels: 48
DB: 2 Gaps: 12
US-09-931-704-4 (1-819) x E88633 (1-371)
```

```
QY 518 CCTGGAGGCTGTATCAGAGTGGCCACAGCTACGCTACGTCGAGTTTCAG----- 474
Db 144 Pro-----ArgGlyProProGlyGlnAlaGlyLeuAspGlyLeuProGlyAla 159
QY 473 CTGTGGCAGCTGACGGTTGAGGCCACGCAAGTAACACAGAGGAGTGTACTGTACGCTCAT 414
Db 160 ProGlyGlnProGlySerAsnGlyGlyAlaGlySerAsnGlyAla----- 174
QY 413 AGTTCTGGGTTCAGCCGACGCTGTCATTGAGGCTTCGCCACACTTCCAAAGTTGACCGTGG 354
Db 175 SerGluGlySerAlaGlyGlyCys-----LysThrCysProAlaGlyPro--- 189
QY 353 CCCTGGGCGAGATTCTCTGCCCCACAGTCGAGGAGGATTGAAGTCAGGCTCGTTGAAAGGGG 294
Db 190 -----ProGlyProProGlyProAlaGlyGlnAlaGlyArgProGly 203
QY 293 GCCCCAGGTAGTTCAGTAGTCCAGCTAAGCTGCGGAGTTGATGCTCCAGGTAGCGGG 234
Db 204 AsnAspGlyGlnProGlyAlaProSerPheGlyGlyValGlyAlaProGlyAlaPro 223
QY 233 TGAGGTTCATAGTTTCTTGATGGAGGGGCCAGG-----CCTGGATCTCTGTGCGA 181
Db 224 GlyProAlaGlyAspAlaGlySerProGlyGlnProGlyAlaProGlyGlnProGlyArg 243
QY 180 TTAAGAGCTGGCACTGCAGGAGGTGCCACAGCACCGTCGTCATAGGCAAGCTAACATCCCC 121
Db 244 ProGlyLysAsnAla-GlnGlyGlySerSerArgProGlyProProGlyProAlaGlyPr 263
QY 120 CACAGAGTCCC-----CTGCTCGAGGTCATGGGGCTGGGGCGCG----- 81
Db 263 oProGlyProProGlyAsnAsnGlyAlaProGlyGlyGlyTyrGlyValGlyProProGl 283
QY 80 -----CGCCGCGCGGGCGGCTCTCTCCAGAGGCTGGCGGAGTGGG 37
Db 283 yProProGlyProSerGlyArgProGlyAla-----ProGlyGlnProGlyProAs 300
QY 36 AGGCGAGCGCGCGGCTCCG 18
Db 300 pGlyGlnProGlyAlaPro 306
RESULT 9
S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N;Alternate names: procollagen alpha 5(IV) chain
N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #text_change 21-Jul-2000
C;Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598;
R;Zhou, J.; Hertz, J. M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ide
n Alport syndrome patient.
A;Reference number: S22917; MUID:92316923; PMID:1352287
A;Accession: S22917
A;Molecule type: mRNA
A;Residues: 1-967 <ZHO>
A;Cross-references: GB:M90464; NID:g180826; PIDN:AAA52046.1; PID:g553234
R;Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A;Title: Structure of the human type IV collagen COL4A5 gene.
A;Reference number: A54365; MUID:94165049; PMID:8120014
A;Accession: A54365
A;Molecule type: DNA
A;Residues: 1-922 <ZH2>
A;Cross-references: GB:U04470; NID:g463378; GB:U04520; NID:g463428; PIDN:AAO27816.1;
R;Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Tryggvason, K.
J. Biol. Chem. 269, 1167-1169, 1993
A;Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited
A;Reference number: A57079; MUID:93361972; PMID:8356449
A;Accession: A57079
A;Molecule type: DNA
A;Residues: 1-27 <ZH4>
```



```

Db 1339 -----GlyMetLysGlyProSerGlyValPro 1347
QY 385 CAATGACAGCGTCGGGTGACCCAGAACTATAGGCGGTACAGTCACCTCTGTGTACTT 444
Db 1348 -----GlySerAlaGlyProGlu-----GlyGluProGlyLeuLeu 1359
QY 445 GCGTGCCTCAACCGTCAGGCTGCCACAGCTGAATCCGACGTAGCCTGGCCCACTTCTG 504
Db 1360 GlyProGlyProGlyLeu----- 1367
QY 505 TACCAGCTCCAGGCGCTGCTGGGAGCATGTCAGGTGTGATGGCAGCGTTGGCTACCC 564
Db 1368 -----ProGlyProSerGlyGlnSerIleIleIleLysGlyAspAla-GlyProPr 1384
QY 565 A-----CTGCCCCAGCGCTCT 579
Db 1384 oGlylleProGlyGlnProGlyLeuLysGlyLeuProGlyProGlnGlyProGlnGlyLe 1404
QY 580 GCCAGGACTGACCGACCTGGGCGCTGCGCCCTGCGCCACAGTGAATTCCTCCAGAAAGAT 639
Db 1404 uProGly-----ProThrGlyProProGlyAspProGlyArgAsnGlyLeuProGlyPh 1422
QY 640 GGATGACTTCTGCTGCTGAGAGCTGTCAGACTGGCTATGGGTTCAGGCCAAGACTT 699
Db 1422 eAspGly-AlaGlyGly---ArgLysGlyAspProGlyLeuProGlyGlnProGlyThr- 1440
QY 700 CAACCGCTTAAGAAGAAGATGCAGCTCCAGCAGCTTCAGTCACCTGCATCTTGGAGGC 759
Db 1441 --ArgGlyLeuAspGlyProProGlyProAspGlyLeuGlnGlyPro----- 1455
QY 760 ACATGTTTCTGACCTCTGACCTTAAACCCACACCTCCAGGCCCACTGAGCTGTGCT 818
Db 1456 -----ProGlyProGlyThrSerSerValala 1465

RESULT 10
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C:Species: Ascaris suum (pig roundworm)
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: S16366
R:Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti
A:Reference number: S16366; MUID:91340768; PMID:1714907
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <JB1>
A:Cross-references: GB:M67507; NID:gl59648; PIDN:AAA18014.1; PID:gl59649
C:Genetics:
A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F:43-1525/Domain: collagenous #status predicted <COL>
F:197-199/Region: cell attachment (R-G-D) motif
F:1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1530-1638/Domain: repeat NC1 #status predicted <NC12>
F:1639-1763/Domain: repeat NC1 #status predicted <NC12>
F:31.34;39.41;536;539/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1593-1599;1702-1709/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 0.00515 Length: 1763
Score: 135.00 Matches: 84
Percent Similarity: 34.13% Conservative: 16
Best Local Similarity: 28.67% Mismatches: 102
Query Match: 9.01% Indels: 91
DB: 2 Gaps: 15

```

```

US-09-931-704-4 (1-819) x S16366 (1-1763)
QY 19 GGAGCGCGGCTCGCCCTCCCACTCCGACGCTCTGGAGAGAGCCGCGCCGCGCGG 78
Db 870 GlyLeuAlaGlyAlaProGlyPheProGlyAlaLysGlyGluProGlyLeu 889
QY 79 CCCGCCCCAG---CCCCATGGACCTCCGAGCAGG-----GGA 114
Db 890 ProGlyLysGlyGluProGlnGlyProProGlyGlnProGlyAlaProGlyPheProGly 909
QY 115 CTCGTGGGGATGTAGCTTGCCT-----ATGCACGGTGTGTG 153
Db 910 GlnLysGlyAspGluGlyLeuProGlyLeuProGlyValSerGlyMetLysGlyAspThr 929
QY 154 GCACCTCCCTCAGTGCAGCTCTTAATCGCACAGAGATCCAGGC---CCTGGGCCCTC 210
Db 930 Gly-LeuProGlyValProGlyLeu-----AlaGlyProProGlyGlnProGly-PheP 947
QY 211 CATCCAGAAAACCTATGACCTCACCCGCTACCTGGAGCATCAACTCCGAGCTTAGCTGG 270
Db 947 roGlyGlnLysGlyGln-Pro---GlyPheProGlyValAlaGlyAlaLysGlyGluAla 965
QY 271 GACCTACTGAACTACCTGGGGCCCTTCAACGAGCTGACTTCAATCCTCCTCGACT 330
Db 966 GlyLeuProGlyLeuProGlyAlaPro----- 974
QY 331 GGGGCGCAGAACTCTGCCCGAGGCCACGCTCAACTTGAAGTGTGGCGAAGCCTCAATGA 390
Db 975 ---GlyGlnLysGlyGluGlnGly----- 981
QY 391 CAGCTCGGCTGACCCAGAACTATAGCGGTACAGTCACCTCCTGTGTACTTGGTGG 450
Db 982 ---LeuAlaGlyLeuProGlyLysGlyAlaProGlyLysGlyAlaProGlyLysGlyAla 1000
QY 451 CCTCAACCGTCAGCTGCCAGCTGAATCCGAGCTAGCTGGC----- 495
Db 1001 ProGlyGlnAspGly-----LeuProGlyLeuProGlyValLysGlyAspArg 1016
QY 496 -----CCACTTCTGTACCAGCTCCAGCGCTCGTGGGAGCATTCAGG 540
Db 1017 GlyPheAsnGlyLeuProGlyGlyLysGlyGluProGlyProAlaAlaArgAspGlyGlu 1036
QY 541 TGTATGCGCAGCTGTGCTACCACTCCGAGCTCTGCCAGGCTCTGCCAGGCTAGAG---CCAGC 597
Db 1037 ---LysGlyGluProGlyLeuProGlyGlnProGly-LeuArgGlyProGlnGlyProPr 1055
QY 598 CTGGGCCCTGGCCCTGCCAGCTGACTTCTCCAGAGATGGATGATCTTCTGGCTGCT 657
Db 1055 oGlyLeuProGly-LeuProGlyLeuLysGlyAspGluGly----- 1068
QY 658 GAAGGAGCTGCAGACCTGGCTATGGCGTTCAGCCAAAGACTTCAACCGCTTAAGAAGAA 717
Db 1069 -----GlnProGlyTyrgly-----AlaProGlyLeuMetGlyG 1080
QY 718 GATCAGCTCCAGCAGCTTCAGTCACCTGCATCCCTGACCTTGGAGGCACATGTTTCTGACCTCT 777
Db 1080 LuLysGlyLeuProGlyLeuProGlyLys----- 1089
QY 778 GACCTTAACCCACACCTCCAGGCCCA 806
Db 1090 --ProGlyArgProGlyAlaProGlyPro 1098

RESULT 11
T15936
hypothetical protein EGAP7.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T15936
R:Miller, N.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid EGAP7.
A:Reference number: Z18433
A:Accession: T15936

```


QY 280 GAACACTACCTGGGCGCCCTTCAACGAGCCTGACTTCAATCCTCCTCGACTGGGGCGAGA 339
 Db GlySerProGlyArgPro-----GlyThrLysGlyGlu 808
 QY 340 AACTCTGCCAGGCGCCAC----- 357
 Db ProGlyAlaProGlyArgValMetThrSerGluGlySerSerThrIleThrValProGly 828
 QY 358 -----GCTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGCGCTGCGGCT 402
 Db ProProGlyProProGlyAlaMetGlyProProGlyProProGlyThrProGlyProAla 848
 QY 403 GACCCAGAA---CTATGAGCGGTACAGTCACT----- 432
 Db GlyProAlaGlyLeuProGlyGlnGlnGlyProArgGlyGluProGlyLeuAlaGlyAsp 868
 QY 433 -----CCTGTGTTACTTGGCTGG----- 450
 Db SerPheLeuSerSerGlySerSerIleSerGluValLeuSerAlaGlnGlyValAspLeu 888
 QY 451 -----CCTCAACCGCTGAGCTGCCACAGCTGAACCTCCGAGCTAGCCTGGCCCACTT 501
 Db ArgGlyProProGlyProProGlyProArgGly-----Pro-----ProGlyProSer 904
 QY 502 CTGTACACGCTCCAGGCGCTGCTGGCAGCATGTCAGGTGTCATGCGCAGCGCTTGGCTA 561
 Db IleProGlyProProGlyProArgGlyPro-----ProGlyGluGly---Val 919
 QY 562 CCCACTGCCCGCCGCTCTGCCAGGG-----ACTGA 591
 Db ProGlyProProGlyPro---ProGlySerPheLeuThrAspSerGluThrPhePheThrGly 939
 QY 592 GCCAGCTGGGCGCCCTGCGCCCTGCCACAGTGACTTCTCCAGAGATGATGACTTCTG 651
 Db ProProGlyProProGlyPro---ProGlyProLysGlyAspGlnGly----- 954
 QY 652 GCTGTGAAGGAGCTGCAGACCTGGCTATGGCTTTCAGCCAGACTTCAACCGGCTTAA 711
 Db -----AspProGlyValProGlyThrProGlyIleSerGlyGlyLeuS 969
 QY 712 GAAGAAGATGCAGCTCCAGCAGCTT-----CAGTCACCCCTGCAGCTTGGAGGCACATGG 765
 Db eHisGlyAlaSerSerSerThrLeuThrMetGlnGlyPro----- 982
 QY 766 TTCTGACCTCTGACCTTAAACCCACACCTCCAGGCCCA 806
 Db -----ProGlyProProGlyProProGlyPro 991
 RESULT 15
 COMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 16-Jun-2000
 C:Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A32003; A31766; S19079
 R:Muthukumar, G.; Blumberg, B.; Kurkinen, M.
 J. Biol. Chem. 264, 6310-6317, 1989
 A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff
 A:Reference number: A33525; MUID:89197932; PMID:2703490
 A:Accession: A33525
 A:Molecule type: mRNA
 A:Residues: 1-1669 <MUT>
 A:Cross-references: EMBL:J04494; NID:G556296; PIDN:AAA50292.1; PID:G556297
 R:Wood, L.; Thieriault, N.; Vogeli, G.
 FEBS Lett. 227, 5-8, 1988
 A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al
 A:Reference number: S01454; MUID:88112221; PMID:3338568
 A:Accession: S01454
 A:Molecule type: mRNA
 A:Residues: 1-185; 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71
 A:Cross-references: EMBL:X06777
 R:Killen, P.D.; Burbello, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988
 A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain

A:Reference number: A28066; MUID:88243724; PMID:3379041
 A:Accession: A28066
 A:Molecule type: mRNA
 A:Residues: 1-129 <K11>
 A:Cross-references: EMBL:J03758; NID:G192869; PIDN:AAA37439.1; PID:G192670
 R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Vos
 Eur. J. Biochem. 147, 217-224, 1985
 A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alph
 A:Reference number: A02864; MUID:85127033; PMID:2578961
 A:Accession: A02864
 A:Molecule type: mRNA
 A:Residues: 1276-1669 <OBE>
 A:Cross-references: EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID:G1333876
 R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
 Gene 43, 301-304, 1986
 A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligod
 A:Reference number: A25636; MUID:86301886; PMID:3755692
 A:Accession: A25636
 A:Molecule type: mRNA
 A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
 A:Cross-references: EMBL:M14042; NID:G192286; PIDN:AAA37342.1; PID:G192287
 A:Note: the authors translated the codon CAG for residue 1374 as Arg
 R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pih
 J. Biol. Chem. 262, 8496-8499, 1987
 A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(I
 A:Reference number: A94680; MUID:87250460; PMID:3597383
 A:Accession: A29301
 A:Molecule type: mRNA
 A:Residues: 1441-1669 <KUR>
 A:Cross-references: EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PID:G387115
 R:Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y.
 J. Biol. Chem. 263, 12310-12314, 1988
 A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequ
 A:Reference number: S19079; MUID:88315019; PMID:2842328
 A:Accession: S19079
 A:Molecule type: DNA
 A:Residues: 1-28 <K12>
 A:Cross-references: EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PID:G46503
 R:Kayes, P.; Wood, L.; Thieriault, N.; Kurkinen, M.; Vogeli, G.
 J. Biol. Chem. 263, 19274-19277, 1988
 A:Title: Head-to-head arrangement of murine type IV collagen genes.
 A:Reference number: A92702; MUID:89066738; PMID:3198626
 A:Accession: A32003
 A:Molecule type: DNA
 A:Residues: 1-28 <KAY>
 A:Cross-references: EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PID:G450449
 R:Burbello, P.D.; Martin, G.R.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
 A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional pr
 A:Reference number: A94220; MUID:89071759; PMID:3200851
 A:Accession: A31766
 A:Molecule type: DNA
 A:Residues: 1-28 <BUR>
 A:Cross-references: EMBL:M23333; NID:G340878; PIDN:AAA51625.1; PID:G535668
 R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
 J. Biol. Chem. 261, 6654-6657, 1986
 A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen ge
 A:Reference number: S19094; MUID:86196099; PMID:3009468
 A:Accession: S19094
 A:Molecule type: DNA
 A:Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>
 A:Cross-references: EMBL:M13027
 R:Schuppan, D.; Timpl, R.; Glanville, R.W.
 FEBS Lett. 115, 297-300, 1980
 A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane
 A:Reference number: S16909; MUID:80246483; PMID:6772473
 A:Accession: S16909
 A:Molecule type: protein
 A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-
 R:Schuppan, D.; Glanville, R.W.; Timpl, R.
 Eur. J. Biochem. 123, 505-512, 1982
 A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a
 A:Reference number: A25991; MUID:82186723; PMID:6804236

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:44:12 ; Search time 10.63 Seconds
(without alignments)

6391.168 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattattaagatttcgcgg.....agccacagtcagctgtgctt 819

Scoring table: BLOSUM62

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cn2 1/USPTO spool/US09931704/runat 27012003 154125 3623/app query.fasta 1.7189
-DB=SwissProt 40 -OPMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09931704 @CGN 1 1 141 @runat 27012003 154125 3623 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	9.7	684	CA39 HUMAN	Q14050 homo sapien
2	142.5	9.7	660	YH11 EBV	P03181 epstein-bar
3	137	9.1	1685	CA54 HUMAN	P29400 homo sapien
4	135	9.0	1763	CA24 ASCSU	P27393 ascaris suu
5	134	9.2	627	SPD2 NEPCCL	P46804 nephila cla
6	133	8.9	1083	T2D3 HUMAN	O00268 homo sapien
7	132	8.8	1669	CA14 MOUSE	P02463 mus musculus
8	131	8.7	317	YQ35 CAEEL	Q09456 caenorhabdi
9	131	8.7	754	CA54 CANFA	Q28247 canis famil
10	130	8.7	660	YH11 EBV	P03181 epstein-bar
11	129.5	8.6	360	CCD2 CAEEL	P35799 caenorhabdi
12	128	8.5	316	CC07 CAEEL	P18832 caenorhabdi
13	127.5	8.7	316	CC12 CAEEL	P20630 caenorhabdi
14	127.5	8.7	316	CC13 CAEEL	P20631 caenorhabdi
15	127.5	8.5	1237	NME3 RAT	Q00961 rattus norv
16	127.5	8.5	1670	CA34 HUMAN	Q01955 homo sapien
17	127	8.5	1255	D1A1 MOUSE	Q08808 mus musculus
18	126.5	8.4	299	CC34 CAEEL	P34687 caenorhabdi

19	126	8.4	1069	1	S24B ARATH	Q9M081 arabidopsis
20	125.5	8.4	415	1	SYN1 CANFA	O62732 canis famil
21	125	8.3	721	1	YK82 MYCTU	Q10690 mycobacteri
22	125	8.3	1362	1	CA21 CHICK	P02467 gallus gall
23	124.5	8.3	494	1	ATF7 HUMAN	P17544 homo sapien
24	124.5	8.3	1806	1	CA1B HUMAN	P12107 homo sapien
25	124	8.3	779	1	CA11 BOVIN	P02453 bos taurus
26	124	8.5	825	1	ICP0 HSV2H	P28284 herpes simp
27	124	8.5	1147	1	MYSB ACACA	P19706 acanthameb
28	124	8.3	1804	1	CA1B MOUSE	Q61245 mus musculu
29	124	8.3	2944	1	CA17 HUMAN	Q02388 homo sapien
30	123.5	8.2	680	1	CA1A MOUSE	Q05306 mus musculu
31	123.5	8.2	743	1	CA18 MOUSE	Q00780 mus musculu
32	123.5	8.2	3530	1	MY15 HUMAN	Q9UKN7 homo sapien
33	123	8.2	680	1	CA1A HUMAN	Q03692 homo sapien
34	122.5	8.2	453	1	CA44 BOVIN	Q29442 bos taurus
35	122	8.1	705	1	SYN1 HUMAN	P17600 homo sapien
36	122	8.1	1460	1	CA11 CANFA	Q9X9J7 canis famil
37	122	8.1	1464	1	CA11 HUMAN	P02452 homo sapien
38	121.5	8.1	296	1	PRP3 MOUSE	P05143 mus musculu
39	121.5	8.1	555	1	GPI CHURE	Q9FDP6 chlamydomon
40	121.5	8.1	744	1	CA18 RABIT	P14282 oryctolagus
41	121.5	8.3	1462	1	CA13 CHICK	P12105 gallus gall
42	121.5	8.1	1838	1	CA15 HUMAN	P20908 homo sapien
43	120.5	8.0	475	1	S3A2 MOUSE	Q62203 mus musculu
44	120.5	8.0	674	1	CA1A CHICK	P08125 gallus gall
45	120.5	8.0	1567	1	FMN2 MOUSE	Q9J104 mus musculu

ALIGNMENTS

RESULT 1

CA39 HUMAN STANDARD: PRT; 684 AA.
ID CA39 HUMAN Q14050; Q9UPE2; Q9H4G9; Q13681;
AC Q14050; Q9UPE2; Q9H4G9; Q13681;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 3(IX) chain precursor.
GN COL9A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=96163887; PubMed=8586434;
RA Lee B., Horton W.A., Olsen B.R., Baker J.R., Mayne R.;
RA "Molecular cloning of the alpha 3 chain of human type IX collagen:
RT linkage of the gene COL9A3 to chromosome 20q13.3";
RL Genomics 30:329-336(1995).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS 563-GLY--PRO-565 DEL AND
RP 564-PRO--GLY-566 DEL.
RX MEDLINE=99357778; PubMed=10428822;
RA Passilita P., Pihlajamaa T., Annunen S., Brewton R.G., Wood B.M.,
RA Johnson C.C., Liu J., Gong Y., Warman M.D., Prockop D.J., Mayne R.,
RA Ala-Kokko L.;
RT "Complete sequence of the 23-kilobase human COL9A3 gene. Detection of
RT Gly-X-Y triplet deletions that represent neutral variants.";
RL J. Biol. Chem. 274:22469-22475(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carter C., Carter N., Clee C.M.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clegg S.,
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,


```

RESULT 2
ID YHL1_EBV          STANDARD;          PRT;          660 AA.
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLFI protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OC NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=60871149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211 (1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; -; NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0B83.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86D1D67A37152A2 CRC64;

Alignment Scores:
Pred. No.: 0.00699 Length: 660
Score: 142.50 Matches: 92
Percent Similarity: 32.43% Conservative: 16
Best Local Similarity: 27.63% Mismatches: 125
Query Match: 9.75% Indels: 100
DB: 1 Gaps: 15

US-09-931-704-4 (1-819) x YHL1_EBV (1-660)
QY 806 TGGGCTGAGGTGGGGGTTAAGGTCAGAGGTCAGAAACCATGTGCTCCCAAGTGCA 747
Db 159 TTPArgArgSerGlyAla-----GlnArgGlyHisProProGlyAla 174
QY 746 GGGTGACTGAAGCTGCTGAGGCTGCATCTTCTTAAGCGGTTGAAGTCTTGGCTG 687
Db 175 GlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThr 194
QY 686 AACGCCATA-----GCCAGGCTGCAGCTCTTTCAGACGCCAGAGTGCATCCA 639
Db 195 ProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaThrProHisPro 214
QY 638 TCTTCTGGAGGAAGTCACCTGGG-----CAGGGCCAG 606
Db 215 GluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGln 234
QY 605 GGGCCAGGCTGGCTCAGTCCTGGCAGAGGCTGGGCGAGTGGTAGCAAGCTGCGCA 546
Db 235 GluProArgLeuProGlnAspLeuAlaAla----- 245
QY 545 TGACACCTGCAATGCTGCCACAGGC----- 519
Db 246 -----GlnArgCysProAlaGlyProProThrArgSerGlyAlaAlaGln 262

```

RN RP SEQUENCE OF 85-1685 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=93244772; PubMed=1363780;
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
 RA Pizzoni G.F., de Marchi M.;
 RT "de novo mutation in the COL4A5 gene converting glycine 325 to
 RT glutamic acid in Alport syndrome.";
 RL Hum. Mol. Genet. 1:127-129(1992).
 RN [13]
 RN RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
 RX MEDLINE=94010948; PubMed=8406498;
 RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
 RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
 RA van Oost B.A., Smeets H.J.M.;
 RT "Identification of four novel mutations in the COL4A5 gene of
 RT patients with Alport syndrome.";
 RL Genomics 17:485-489(1993).
 RN [14]
 RN RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872
 RX AND C-1241.
 RX MEDLINE=95322976; PubMed=7599631;
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
 RT "Detection of 12 novel mutations in the collagenous domain of the
 RT COL4A5 gene in Alport syndrome patients.";
 RL Hum. Mutat. 5:197-204(1995).
 RN [15]
 RN RP VARIANTS AS ARG-1649.
 RX MEDLINE=96213750; PubMed=8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 RA Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 RT common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RN RP VARIANTS AS
 RX MEDLINE=96213754; PubMed=8651296;
 RA Turco A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RA Turco A., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 RT exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RN RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
 RX MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacaeseier D.,
 RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 RT syndrome.";
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RN RP VARIANTS AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RA Tverskaya S., Bobryniina V., Tsalykova F., Ignatova M.,
 RA Krasnopolskaya X., Evgrafov O.;
 RT "Substitution of A1498D in noncollagen domain of a5(IV) collagen
 RT chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RN [19]
 RN RP VARIANTS AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RT syndrome and COL4A5 R1677Q.";
 RL Hum. Genet. 99:681-684(1997).
 RN [20]
 RN RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
 RX MEDLINE=98112435; PubMed=9452056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
 RA

RN RP SEQUENCE OF 924-1685 FROM N.A.
 RX MEDLINE=91169491; PubMed=2004755;
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
 RT "Characterization of the 3' half of the human type IV collagen alpha
 RT 5 gene that is affected in the Alport syndrome.";
 RL Genomics 9:1-9(1991).
 RN [5]
 RN RP SEQUENCE OF 914-1685 FROM N.A.
 RX MEDLINE=90160375; PubMed=1689491;
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
 RA Tryggvason K.;
 RT "Identification of a distinct type IV collagen alpha chain with
 RT restricted kidney distribution and assignment of its gene to the
 RT locus of X chromosome-linked Alport syndrome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
 RN [6]
 RN RP SEQUENCE OF 1442-1471 FROM N.A.
 RX MEDLINE=90252791; PubMed=2339699;
 RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,
 RA Sheer D., Solomon E., Pihlajaniemi T.;
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
 RT to the region of the X chromosome containing the Alport syndrome
 RT locus.";
 RL Am. J. Hum. Genet. 46:1024-1033(1990).
 RN [7]
 RN RP SEQUENCE OF 1-20 FROM N.A.
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
 RA Marynen P.;
 RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RN RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).
 RX MEDLINE=94133540; PubMed=8301933;
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
 RA Cassiman J.-J., Marynen P.;
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood
 RT cells: a complex mutation in the COL4A5 gene of an Alport patient
 RT deletes a complex mutation in the COL4A5 gene of an Alport patient
 RL Kidney Int. 44:1316-1321(1993).
 RN [9]
 RN RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 RN [10]
 RN RP VARIANT AS SER-1564.
 RX MEDLINE=91169492; PubMed=1672282;
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
 RA Tryggvason K.;
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
 RT conserved cysteine to serine in Alport syndrome.";
 RL Genomics 9:10-18(1991).
 RN [11]
 RN RP VARIANT AS ARG-325.
 RX MEDLINE=92303559; PubMed=1376965;
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
 RA Tryggvason K., Gubler M.-C., Antignac C.;
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5
 RT (IV) chain associated with X-linked Alport syndrome: characterization
 RT of the mutation by direct sequencing of PCR-amplified lymphoblast
 RT cDNA fragments.";
 RL Am. J. Hum. Genet. 51:135-142(1992).
 RN [12]

DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001442; ProcollagenC4.
 DR Pfam; PF01391; Collagen; 25.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD000007; Collagen; 3.
 DR ProDom; PD003923; ProcollagenC4; 2.
 DR SMART; SM00111; C4; 2.
 KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 KW Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 1763
 FT DOMAIN 27 42
 FT DOMAIN 27 42
 FT DOMAIN 43 1529
 FT DOMAIN 1530 1763
 FT DISULFID 1548 1637
 FT DISULFID 1581 1634
 FT DISULFID 1593 1599
 FT DISULFID 1656 1752
 FT DISULFID 1690 1749
 FT DISULFID 1702 1709
 FT CARBOHYD 126 126
 FT CARBOHYD 249 249
 FT VARSPLIC 230 266
 FT FT
 FT FT
 FT FT
 FT FT
 SQ SEQUENCE 1763 AA; 169526 MW; 304F528BC06A8E0D CRC64;
 Alignment Scores:
 Pred. No.: 0.0237 Length: 1763
 Score: 135.00 Matches: 84
 Percent Similarity: 34.13% Conservative: 16
 Best Local Similarity: 28.67% Mismatches: 102
 Query Match: 9.01% Indels: 91
 DB: 1 Gaps: 15
 US-09-931-704-4 (1-819) x CA24_ASCSU (1-1763)
 QY 19 GGAGCGGGCTCGCCTCCCACTCGCCAGCTCTGGAGAGAGCGCGCGCGG 78
 DB 870 GlyLeuAlaGlyAlaProGlyPheProGlyAlaGlyGlyGluProGlyLeuProGlyLeu 889
 QY 79 CCGCGCCCCAG---CCCATGACCTCCGAGCAGG---GGA 114
 DB 890 ProGlyLysGlyGluProGlnGlyProGlyGlnProGlyAlaProGlyPheProGly 909
 QY 115 CTCGTGGGAGTGTAGTTGCTT-----ATGACGGTGTGTG 153
 DB 910 GlnLysGlyAspGluGlyLeuProGlyLeuProGlyValSerGlyMetLysGlyAspThr 929
 QY 154 GCACCTCCCTGCAGTCCAGCTCTTAATCGACAGAGATCCAGG---CCTGGCCCTC 210
 DB 930 Gly-LeuProGlyValProGlyLeu-----AlaGlyProGlyGlnProGly-PheP 947
 QY 211 CATCCAGAAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGACCTTAGCTGG 270
 DB 947 roGlyGlnLysGlyGln-Pro---GlyPheProGlyValAlaGlyAlaLysGlyGluAla 965
 QY 271 GACCTACCTGAACCTACTGGGGCCCTTCAACGACCTGACTTCACTCTCTCCTCACT 330
 DB 966 GlyLeuProGlyLeuProGlyAlaPro----- 974
 QY 331 GGGGCGAGAACTCTGCCAGGGCCACGGTCAACTTGGAGTGTGGAGACCTCAATGA 390
 DB 975 ---GlyGlnLysGlyGluGlnGly----- 981
 QY 391 CAGGCTCGGCTGACCCAGAACTATGAGCGGTACAGTCACTCTGTGTTACTTGGTGG 450
 DB 982 ---LeuAlaGlyLeuProGlyIleProGlyMetLysGlyAlaProGlyIleProGlyAla 1000
 QY 451 CCTCAACCGTCAGGCTGCCAGCTGAACCTCCGACGTAGCTGGC----- 495
 DB 1001 ProGlyGlnAspGly-----LeuProGlyLeuProGlyValLysGlyAspArg 1016

QY 496 -----CCACTTCTGTACCAAGCTCCAGGCGCTGCTGGGACGACATTGCAGG 540
 DB 1017 GlyPheAsnGlyLeuProGlyGlyLysGlyGluProGlyProAlaAlaArgAspGlyGlu 1036
 QY 541 TGTATGGGACGCTGGCTTACCACCTGCCAGCCTCTGCCAGGACTGAG---CCAGC 597
 DB 1037 ---LysGlyGluProGlyLeuProGlyGlnProGly-LeuArgGlyProGlnGlyPro 1055
 QY 598 CTGGGCGGCTGGCCCTGCCACAGTACTCTCCAGAAGATGATGACTTCTGGCTGCT 657
 DB 1055 oGlyLeuProGly-LeuProGlyLeuLysGlyAspGlyGly----- 1068
 QY 658 GAAGGAGCTGCAGACCTGGCTTATGGCTTTCAGCCAGGACTTCAACCGCTTAAGAAGAA 717
 DB 1069 -----GlnProGlyTyrgly-----AlaProGlyLeuMetGlyG 1080
 QY 718 GATGACGCTCCAGAGCTTCAGTCACCTGTCACCTGGAGCAGCATGTTTCTGACCTCT 777
 DB 1080 lulyGlyLeuProGlyLeuProGlyLys----- 1089
 QY 778 GACCTTAACCCACACACCTCCAGGCCCCA 806
 DB 1090 --ProGlyArgProGlyAlaProGlyPro 1098
 RESULT 5
 SPD2_NEPCL STANDARD; PRT; 627 AA.
 AC P46804;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Spidroin 2 (Dragline silk fibroin 2) (Fragment).
 OS Nephila clavipes (Orb spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
 OX NCBI_TaxID:6915;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92406876; PubMed=1527052;
 RA Hinman M.B., Lewis R.V.;
 RT "Isolation of a clone encoding a second dragline silk fibroin."
 RL J. Biol. Chem. 267:19320-19324 (1992).
 CC -!- FUNCTION: Spiders major ampullate silk possesses unique
 CC characteristics of strength and elasticity. Fibroin consists of
 CC pseudocristalline regions of antiparallel beta-sheet interspersed
 CC with elastic amorphous segments.
 CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- DOMAIN: Highly repetitive protein characterized by regions of
 CC polyalanine and glycine-rich repeating units.
 CC -!- SIMILARITY: BELONGS TO THE SILK FIBROIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M92913; AAA29381.1; -.
 CC Silk; Repeat.
 KW NON TER 1 1
 FT DOMAIN 1 530 15 APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1 36 1.
 FT REPEAT 37 79 2.
 FT REPEAT 80 121 3.
 FT REPEAT 122 172 4.
 FT REPEAT 173 213 5.
 FT REPEAT 214 252 6.
 FT REPEAT 253 283 7.

15-JUL-1998 (Rel. 36, Last sequence update)
DT
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DR Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
TX (TAFII135) (TAFII-130) (TAFII130).
GN
DE TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP
RX MEDLINE=7336072; PubMed=9192867;
MEDLINE=97336072; PubMed=9192867;
EN
RA Mengus G., May M., Carre L., Chambon P., Davidson I.;
RA "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT mammalian cells.";
RL Genes Dev. 11:1381-1395(1997).
RN [2]
RP
RX SEQUENCE FROM N.A.
MEDLINE=21638749; PubMed=11780052;
TX DeLoukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond E., Hunt A.R., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckie S., Hurst J.L., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehmalaaho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.I., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillips B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaubin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP
RX SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=97098442; PubMed=8942982;
TX Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RA "Molecular cloning and analysis of two subunits of the human TFIID
RT complex: hTAFII130 and hTAFII100."
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: Y11354; CAA72189.1; -.
DR EMBL: AL137077; CAC36006.1; -.
DN


```

DR EMBL; AL109911; CAC22312.2; -.
DR EMBL; U75308; AAC50901.1; -.
DR TRANSFAC; T02328; -.
DR Genew; HGNC:11537; TAF4.
DR MIM; 601796; -.
DR InterPro; IPR003894; TAF_hom.
DR SMART; SMO0549; TAFH; 1.
KW Transcription regulation; Nuclear protein.
FT DOMAIN 39 42 POLY-HIS.
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 98 101 POLY-GLY.
FT DOMAIN 142 148 POLY-ALA.
FT DOMAIN 268 275 POLY-PRO.
FT DOMAIN 331 337 POLY-ALA.
FT DOMAIN 680 683 POLY-PRO.
FT DOMAIN 808 813 POLY-ALA.
FT DOMAIN 828 831 POLY-ASP.
FT DOMAIN 105 117 PGPPSPRRPLVPA -> GRGLLQRRGRES
FT CONFLICT 136 136 A -> S (IN REF. 2).
FT CONFLICT 185 185 G -> GPG (IN REF. 2).
FT CONFLICT 233 264 MISSING (IN REF. 3).
FT CONFLICT 293 293 P -> L (IN REF. 3).
SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Alignment Scores:
Pred. No.: 0.0326 Length: 1083
Score: 133.00 Matches: 81
Percent Similarity: 35.45% Conservative: 25
Best Local Similarity: 27.09% Mismatches: 123
Query Match: 8.88% Indels: 70
DB: 1 Gaps: 14

US-09-931-704-4 (1-819) x T2D3_HUMAN (1-1083)
Qy 19 GGAGCGGGCTGCGCCTCCCACTCGCCAGCTCTGGGAGAGAGCGCGCCGCGCGG 78
Db 81 GlyAlaProGlyAlaAlaProGluProProProAlaGlyArg--AlaArgProGlyGly 99
Qy 79 CCGCGCCCGCCAGCCCATGAGCTCCGAGC-----AGGGAGCTCTGGGGATGTAGC 132
Db 100 GlyGlyProGlnArgProGlyProProSerProArgProLeuValProAlaGlyPro 119
Qy 133 TTGCCTATGACAGCGTCTGTCGGACCTCCCTGCGAGTGCAGTCTTAATCGCACAGAGA 192
Db 120 AlaProProAlaAlaLysLeuArgProProProGluGlySerAlaGlyAlaCysAla--- 138
Qy 193 TCCAGGCGCTGCGCCCTCCATCCAGAAAACCTATGACCTCCCGCTACTCGGAGCATCA 252
Db 139 -----ProValProAlaAlaAla 144
Qy 253 ACT-----CCGCGAGCTTAGTGGACCTACCTGAACCTAGCTGGGCGCCCTTT 300
Db 145 AlaValAlaAlaGlyProGluProAlaProAlaGlyProAlaLysProAlaGlyProAla 164
Qy 301 CAACGAGCGCTGACTTCAATCTCTCGACTGGG----- 333
Db 165 AlaLeuAlaAlaArgAlaGlyProGlyProGlyProGlyProGlyProGlyPro 184
Qy 334 GGCAGAACTTGGCCAGGGCCACGGTCAACTT-----GGAAGTGTGGCGAGCCTCAA 387
Db 185 GlyLysProAlaGlyProGlyAlaAlaGlnThrLeuAsnGlySerAlaAlaLeuLeuAsn 204
Qy 388 TGACAGCGCTGCGCTGACCCAGCACTATGAGGGGTACAGTCA---CCTCCTGTGTACTT 444
Db 205 SerHisHisAlaAlaAlaProAlaAlaValSerLeuValAsnAsnGlyProAlaAlaLeuLeu 224
Qy 445 GCGTGGCCTCAACCGTCAGCGCTGCCACAGCTGAACCTCCGAGTACGCTGGCCCACTTCTG 504
Db 225 ProLeuProLysProAlaAla-----ProGlyThrValIle 236
Qy 505 TACAGCCTCCA-----GGCGCTGCTGGCGAGCATTCGACGTTGTCATGGC 549

```

```

Db 237 GlnThrProProPheValGlyAlaAlaProProAlaProAlaAlaProSerProPro 256
Qy 550 GAGCGTTGGCTACCGCTGCCAGCTCTGCTCCAGGGAGCTGAGCCAGCGCTGGCCCTGG 609
Db 257 AlaAlaProAlaProAlaAlaProAlaAlaProPro-----ProProPro 272
Qy 610 CCTGCCCCACAGTGACTT-----CCTCCA-----GAAGATGGATGACTTCTGGCTGCT 657
Db 273 ProAlaProAlaThrLeuAlaArgProProGlyHisProAlaGlyProProThrAlaAla 292
Qy 658 GAAGGAGCTGCAGACCTGGCTATGGCTTGCAGCCCAAGAGCTTCAACCGCTTAAGAAGA 717
Db 293 ProAlaValProProProAlaAlaAlaAlaGlnAsnGlyGly----- 305
Qy 718 GATGCGACCTCCAGCAGCTTCCAGTCCACCTGTCACCTGGAGGCACATGTTTGTGACCTCT 777
Db 306 SerAlaGlyAlaAlaProAlaProAlaProAlaAlaGlyGlyProAlaGlyValSerGly 325
Qy 778 GACCTTTAACCC-----CCACACCTCCAGGCCAGTCA 810
Db 326 GlnProGlyProGlyAlaAlaAlaAlaProAlaProGlyValLysAlaGluSer 344

RESULT 7
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DC 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89137932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glangville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen."

```

J. Biol. Chem. 262:8496-8499 (1987).
 [6]
 RN PARTIAL SEQUENCE FROM N.A.
 RP MEDLINE=86196099; PubMed=3009468;
 RA Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
 collagen genes.";
 RL J. Biol. Chem. 261:6654-6657 (1986).
 [7]
 RN SEQUENCE OF 1-28 FROM N.A.
 RP MEDLINE=89066738; PubMed=3198626;
 RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
 RT "Head-to-head arrangement of murine type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277 (1988).
 [8]
 RN SEQUENCE OF 1-28 FROM N.A.
 RP MEDLINE=89071759; PubMed=3200851;
 RA Burdello P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
 bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682 (1988).
 [9]
 RN SEQUENCE OF 1-129 FROM N.A.
 RP MEDLINE=88243724; PubMed=3379041;
 RA Killen P.D., Burdello P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
 collagen chain and the corresponding region of the gene.";
 RL J. Biol. Chem. 263:8706-8709 (1988).
 CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; J03758; AAA37439.1; -
 CC EMBL; M23333; AAA51625.1; -
 CC EMBL; J04594; AAA50292.1; -
 CC EMBL; X06777; CAA29946.1; -
 CC EMBL; X02201; CAA26132.1; -
 CC EMBL; M15832; AAA37340.1; -
 CC EMBL; M14042; AAA37342.1; -
 CC EMBL; M12879; AAA37343.1; -
 CC EMBL; M13024; -; NOT_ANNOTATED_CDS.
 CC EMBL; M13025; -; NOT_ANNOTATED_CDS.
 CC EMBL; M13026; AAA37344.1; -
 CC EMBL; M13027; AAA37345.1; -
 CC EMBL; M13043; AAA37346.1; -
 CC EMBL; J04448; AAA37437.1; -
 CC PIR; A33525; CGMS48.
 CC MGD; MGI:88454; Col4a1.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001442; ProcollagenC4.

DR Pfam; PF01391; Collagen; 23.
 DR Pfam; PF01413; C4; 2.
 DR ProDom; PD000007; Collagen; 4.
 DR ProDom; PD003923; ProcollagenC4; 2.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 FT DISULFID 1505 1511 BY SIMILARITY.
 FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
 FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
 FT DISULFID 1616 1622 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 26 26 A -> P (IN REF. 2).
 FT CONFLICT 186 186 S -> L (IN REF. 2).
 FT CONFLICT 319 319 Q -> S (IN REF. 2).
 FT CONFLICT 369 369 Q -> L (IN REF. 2).
 FT CONFLICT 403 403 L -> F (IN REF. 2).
 FT CONFLICT 481 481 P -> L (IN REF. 2).
 FT CONFLICT 493 493 Q -> H (IN REF. 2).
 FT CONFLICT 712 712 S -> I (IN REF. 2).
 FT CONFLICT 813 813 E -> Q (IN REF. 2).
 FT CONFLICT 982 982 V -> H (IN REF. 2).
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).
 SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058B9 CRC64;
 Alignment Scores:
 Pred. No.: 0.0384 Length: 1669
 Score: 132.00 Matches: 75
 Percent Similarity: 34.26% Conservative: 11
 Best local Similarity: 29.88% Mismatches: 91
 Query Match: 8.81% Indels: 75
 DB: 1 Gaps: 15
 US-09-931-704-4 (1-819) x CA14_MOUSE (1-1669)
 QY 13 TTCGCCGAGCGCGGCTGCCCTCCACTCCGCGCAGCCCTCTGGGAGAGGAGCGCGGCC 72
 Db 1193 PheProGlyLeuAlaGlySerProGlyIleProGlyValLysGlyGluGlnGlyPheMet 1212
 QY 73 GCGCGCGCGCGCGCCGAG-----CCCCAT 96
 Db 1213 GlyProProGlyProGlnGlyGlnProGlyLeuProGlyThrProGlyHisProValGlu 1232
 QY 97 GGACCTCCG---AGCAGGGGACTCGTGGGGGATGTTAGCTTCCTATGCAC---GGTGCT 150
 Db 1233 GlyProLysGlyAspArgGlyProGlnGlnProGlyLeuProGlyHisProGlyPro 1252
 QY 151 GTGGCACTTCCTCGCAGTCCAGCTCTTAAT---CGCAGCAGGAGATCCAGGCCCC----- 201
 Db 1253 MetGlyPro-ProGlyPheProGlyIleAsnGlyProLysGlyAspLysGlyAsnGlnGln 1272
 QY 202 -TGGCCCTCCATCCAGAAACCTATGACTCACC----- 237
 Db 1272 yTrpProGlyAlaProGlyVal-----ProGlyProLysGlyAspProGlyPheGlnGln 1290
 QY 238 -----CTACCTGGAGCATCAACTCCGAGCTTAGCTGGGAGCTACC 278
 Db 1290 yMetProGlyIleGlySerProGlyIleThrGlySerLysGlyAspMetGlyLeuPr 1310
 QY 279 TGAACCTACTGGGGCCCTTTCAACAGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAG 338
 Db 1310 oGlyValProGlyPheGlnGlyGlnLys-----GlyLe 1321
 QY 339 AAACCTGCCCGAGGC---CAGGTCAACTTGGAGTGTGGGAGAGCCTCAATGACAGGC 395
 Db 1321 uProGlyLeuGlnGlyValLysGlyAspGlnGlnGlyAsp-----GlnGln 1335

[illegible]

FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;
Alignment Scores:
Pred. No.: 0.0527 Length: 660
Score: 130.00 Matches: 91
Percent Similarity: 31.27% Conservative: 15
Best Local Similarity: 26.84% Mismatches: 97
Query Match: 8.68% Indels: 136
DB: 1 Gaps: 20
US-09-931-704-4 (1-819) x YHLL_EBV (1-660)
QY 19 GGAGCGGGCTCGCCCTCCCTCCAGCTCGGAGGAGGAGCGCGCGCGCGG 78
Db GlyAlaGlyGlnArg---ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAla 191
QY 79 CCGGCGC-----CCCGAGCCCATGGA----- 99
Db ProGlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSerGlyAlaThr 211
QY 100 -----CCTCCG 105
Db ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProPro 231
QY 106 AGCAGGGGACTCGTGGGATGTAGCTTGCCTATGACGGTCTGTGGCACCTCCCTGC 165
Db GluArg-----GlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCys 248
QY 166 -----AGTCCAGCTCTTAATCGCACGAGAGA--- 192
Db ProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgPro 268
QY 193 -----TCCAGGCGCTGGCCCTCCATCCA 216
Db ProGlyCysProArgSerAlaArgAsnProGlyCysProArgThrProArgArgSer 288
QY 217 GAAACCTATGACCTCACCGCTACTCGGAGCA-----TCAACTCCGCGAGCTT 264
Db GlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArgProSerGlyPro----- 306
QY 265 AGCTGGGACTACCTGAACCTACCTGGGGCCCTTTCAACGAGCGCTGACTTCAATCCTCC 324
Db ThrGlyGlyArgProAlaAlaProGlyAlaPro----- 317
QY 325 TCGACTGGGGGAGAACTCTGCCAGGCGCACGTCACCTTGAAGTGTGGGAGCGCT 384
Db -----GlyThrProAlaAlaProGlyProGly-----GlyGlyAlaAlaValPro 332
QY 385 CAATGACAGGCTCGGCTGACCCAGCACTA---TGAGGCGGTACAGTCACCTCCTGTGTTA 441
Db Ser---GlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAlaAla 351
QY 442 CTTCGTGGCT-----CAACCGTCAGGCTGCCACAG-----CTGAACCTCCGAGC 486
Db AlaArgLeuProProGluArgGlnGluPro--ArgLeuProGlnAspLeuAlaAlaAlaG 371
QY 487 TAGCTGGCCCACTCTGTACACGCTCCAGGCGCTGCTGGGAGCATTCGAGGTGCAT 546
Db 371 lnArgCysProAlaGlyProProProThrArgSer---GlyAlaAlaGlnArgThrH 390
QY 547 GCGGAGCTGGCTACCTACCTGCCCCAGCCTCTGCCAGGAGCTGAGCGCTGGGCCCC 606
Db 390 isArgArg---ProProGlyCysPro----- 397
QY 607 TGCCCTGCCCCAGTGAAGTCTTCCAGAGATGGATGATCTTCGCTGTGTAAGAGCT 666
Db 398 -----ArgSerAlaArgAsnProGlyCysProArgThrT 409
QY 667 GCAGA-----CCTGGCTATGGCTTCAGGCA 693
Db 409 rpArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgPro 429

QY 694 GGACTTCAACCGGCTTAAGAGAAGATGCAGCTTCCAGAGCTTCAAGTCCAGCTTCACTT 753
Db erGlyProThrGlyGlyArgPro-----AlaAlaProGlyAla- 441
QY 754 GGAGGCACATGGTTCTGACCTTAACCCCCACACCTCCAGGCCCA 806
Db 442 -----ProGlyThrProAlaAlaProGlyPro 450
RESULT 11
CCD2_CABEL STANDARD; PRT; 360 AA.
ID CCD2_CABEL AC F35799; Q22477; AC F35799; Q22477; Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cuticle collagen dpy-2 precursor.
GN dpy-2 OR T14B4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94060446; PubMed=8241567;
RA Levy A.D., Yang J., Kramer J.M.;
RT "Molecular and genetic analyses of the Caenorhabditis elegans dpy-2
RT and dpy-10 collagen genes: a variety of molecular alterations affect
RT organismal morphology.";
RL Mol. Biol. Cell 4:803-817 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Pavello T.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -1- DISEASE: MUTATIONS IN DPY-2 AFFECTS THE BODY SHAPE.
CC -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; L12706; AAA17398.2; --
CC EMBL; U50191; AAK31557.1; --
CC WormPep; T14B4.6; CE04953.
CC InterPro; IPR002486; Col_cuticle_N.
CC InterPro; IPR000087; Collagen.
CC Pfam; PF01391; Collagen; 2.
CC Pfam; PF01484; Col_cuticle_N; 1.
CC Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
CC Signal.
CC SIGNAL 1 ?
FT CHAIN ? 360
FT DOMAIN 123 152
FT DOMAIN 174 230
FT DOMAIN 238 303
FT VARIANT 129 129
FT VARIANT 183 183
FT VARIANT 247 247
FT VARIANT 253 253
FT CONFLICT 355 360
POTENTIAL.
CUTICLE COLLAGEN DPY-2.
TRIPLE-HELICAL REGION.
TRIPLE-HELICAL REGION.
TRIPLE-HELICAL REGION.
G -> E (IN DPY2 (SC38)).
G -> R (IN DPY2 (E8)).
G -> R (IN DPY2 (E489)).
G -> R (IN DPY2 (Q292)).
PRIRKW -> DQYNGESVSNHKNNGYVHLRFTQ (IN
REF. 2).

Qy 169 GCCAGCTCTTAATCGACAGGAGATCCAGGCGCTGCGCCCTCATCCAGAAACCTATGA 228
 Db 171 -----ProglyPro----- 173
 Qy 229 COTACCCGCTACTGGAGCATCACTCCGAGCTTAGCTGGGACCTACCTGAACCTACCT 288
 Db 174 ProglyProAlaGlySerAlaGlySerProglyGlnAla---GlyAlaProglyAsnPro 192
 Qy 289 GGGGCCCCCTTTCAACGAGCTGACTTCAATCTCTCGACTGGGGGCGAGAACTCTGCC 348
 Db 193 GlySerPro-----GlySerGlyGln 200
 Qy 349 CAGGCGCCAGCTCACTTGGAACTGTGGCAAGCTCAATGACAGGCTGGGCTGACCCA 408
 Db 201 ArgGlyArgGlyLeuProglyProSerGlyAlaPro----- 212
 Qy 409 GAACTATGAGCGGTACAGTCACCTCTGTGTTACTTGGCTGCGCTCAACCGTCAGGCTGC 468
 Db 213 -----GlyProGlnGlyProProgly-----AlaProglyGlnProglySerGly 227
 Qy 469 CACAGCTGAACCTCCGAGCTAGCTGGCCCACTTCTGTACAGGCTCCAGGCTGCTGGG 528
 Db 228 AsnAla-----ProglyProAla-----GlyProProglyProAlaGly 240
 Qy 529 CAG-----CATTCAGGTGTCATGGCGAGCTTGGCTTACCCACTGCCCCA 573
 Db 241 ProAsnGlyGlnProglyHis-ProglyGlnAspGlyGlnProglyAlaProglyAsnAs 260
 Qy 574 GCCTCTGCCAGGACTGAGCAGCTGGGCGCCCTGGCCCTGCCACACAGTACTTCTCCCA 633
 Db 260 pGlyThrProglySerAspAlaAlaGlyCysProCysProThrArgSerValLeuAr 280
 Qy 634 G 634
 Db 280 g 280
 RESULT 13
 ID CC12_CAEEL STANDARD; PRT; 316 AA.
 AC P20630;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cuticle collagen 12 precursor.
 GN COL-12 OR F15H10.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=90172409; PubMed=1689778;
 RA Park Y.-S., Kramer J.M.;
 RT "Tandemly duplicated Caenorhabditis elegans collagen genes differ in
 RT their modes of splicing";
 RL J. Mol. Biol. 211:395-406(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=90172409; PubMed=1689778;
 RA Berke M.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 CC LINKS.
 CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; X51622; CAA35954.1; -;
 DR EMBL; Z73972; CAA98257.1; -;
 DR PIR; S08169; S08169.
 DR WormPep; F15H10.1; CE05638.
 DR InterPro; IPR002486; Col cuticle_N.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF01484; Col cuticle_N; 1.
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;
 KW Signal.
 FT SIGNAL 1 36 POTENTIAL.
 FT CHAIN 37 316 CUTICLE COLLAGEN 12.
 FT DOMAIN 128 157 TRIPLE-HELICAL REGION.
 FT DOMAIN 176 202 TRIPLE-HELICAL REGION.
 FT DOMAIN 206 235 TRIPLE-HELICAL REGION.
 FT DOMAIN 240 265 TRIPLE-HELICAL REGION.
 FT DOMAIN 269 304 TRIPLE-HELICAL REGION.
 SQ SEQUENCE 316 AA; 30098 MW; 6CA81FF94706D42E CRC64;
 Alignment Scores:
 Pred. No.: 0.0786 Length: 316
 Score: 127.50 Matches: 64
 Percent Similarity: 33.19% Conservative: 12
 Best Local Similarity: 27.95% Mismatches: 66
 Query Match: 8.72% Indels: 87
 DB: 1 Gaps: 13
 US-09-931-704-4 (1-819) x CC12_CAEEL (1-316)
 Qy 638 TCTTCTGAGGAAGTCACTGTGGCGAGGCGCCAGGCCCGCCAGG---CTGGCTCAGTCCCTG 582
 Db 117 SerSerGlyGlySerCysCysSerGlyGlyAlaAlaGlyProAlaGlySerPro 136
 Qy 581 GCAGAGCTGGGCGAGTGGGTAGCCCAAGCGTCGCCA----- 546
 Db 137 GlyGlnAspGlyAlaProglyAsnAspGlyAlaProglyAlaProglyAsnProglyGln 156
 Qy 545 -----TGACACCTGCATATGC-----TGCCCGAGGAGGC 519
 Db 157 AspAlaSerGluAspGlnThrAlaGlyProAspSerPheCysPheAspCysProAlaGly 176
 Qy 518 CTGGAGGCTGGTACAGAGTGGGCGAGGCTACGTCGGAGTTGAGTGGCGAGCTGAC 459
 Db 177 Pro-----ProglyProSerGlyAlaPro----- 184
 Qy 458 GGTGAGGCCACGCAAGTAACACAGGAGGTGACTGTACGCCTCATAGTTCTGGGTGAGCC 399
 Db 185 -----GlyGlnLysGlyProSerGlyAlaProglyAlaProglyGlnSerGlyGlyAla 202
 Qy 398 GCAGCTCTCATTTAGGCTTCCGCCACATTCACAGTTGACCGTGGCGCTGGCGAGAGTTT 339
 Db 203 Ala-----LeuProglyProProglyPro----- 210
 Qy 338 CTGCCCCCAGTCGAGGAGGATTGAAGTCAGGCTCGTTGAAGGGGGCCCGAGGTAGTTCA 279
 Db 211 -----AlaGlyProProglyProAla 217
 Qy 278 GGTAGGTCCAGCTAAGTTCGCGAGTTGATGCTCCAGGTCAGGCGGTGAGGTCTAGGTTT 219
 Db 218 GlyGlnProglySerAsnGlyAsnAlaGlyAlaProglyAla----- 231
 Qy 218 TCTGGATGAGGGCGAGGCGCTCGATCTCTGTGCGATTAGAGCTGGCAGCTGCAGGGA 159
 Db 232 -----ProglyGlnValValAspVal-----Pro 239
 Qy 158 GGTGCCACAGCAGCTGTCATAGGCAAGCTAACATCCCCACAGAGTCCCTGCTCGGAGGT 99

Db 240 GlyThrProGlyProAla-----GlyProProGlySerPro-----Gly 252

QY 98 CCATGGCGTGGGGCGGGCGGGCGGGCGGGCTCTCTCCACAGAGGTGGCGGAGTG 39

Db 253 ProAlaGly---AlaProGlyGlnProGlyGlnAlaGlySerSerGlnProGlyGlyPro 271

QY 38 GGA-----GGCGAGCGGGCGGCTCCG 18

Db 272 GlyProGlnGlyAspAlaGlyAlaPro 280

RESULT 14

CC13_CABEL STANDARD; PRT; 316 AA.

AC P20631;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Cuticle collagen 13 precursor.

GN COL-13 OR F15H10.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

ON NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2;

RX MEDLINE=90172409; PubMed=1689778;

RA Park Y.-S., Kramer J.M.;

RT "Tandemly duplicated Caenorhabditis elegans collagen genes differ in

RT their modes of splicing.";

RL J. Mol. Biol. 211:395-406(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2;

RA Berks M.;

RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE

CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A

CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.

CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE

CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-

CC LINKS.

CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; X51623; CAA35955.1; -;

DR EMBL; Z73972; CAA98258.1; -;

DR PIR; S08170; S08170.

DR WormPep; F15H10.2; CE05639.

DR InterPro; IPR002486; Col_cuticle_N.

DR InterPro; IPR000087; Collagen.

DR Pfam; PF01391; Collagen; 3.

DR Pfam; PF01484; Col_cuticle_N; 1.

KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;

KW Signal.

FT SIGNAL. 1 36

FT CHAIN 37 316

FT DOMAIN 128 157

FT DOMAIN 176 202

FT DOMAIN 206 235

FT DOMAIN 240 266

FT DOMAIN 269 304

FT DOMAIN 316 316

SQ SEQUENCE 316 AA; 30100 MW; 00C6D08FBC4701AF CRC64;

Alignment Scores:

Pred. No.: 0.0786 Length: 316

Score: 127.50 Matches: 64

Percent Similarity: 33.19% Conservative: 12

Best Local Similarity: 27.95% Mismatches: 66

Query Match: 8.72% Indels: 87

DB: 1 Gaps: 13

US-09-931-704-4 (1-819) x CC13_CABEL (1-316)

QY 638 TCCTTGGAGGAAGTCACTGTGGCGAGGCCAGGGGCCAGG---CTGGCTCAGTCCCTGTG 582

Db 117 SerSerGlyGlySerCysSerCysGlySerGlyAlaAlaGlyProAlaGlySerPro 136

QY 581 GCAGAGGCTGGGCGAGTGGGTAGCAAGCTCGCCA----- 546

Db 137 GlyGlnAspGlyAlaProGlyAsnAspGlyAlaProGlyAlaProGlyAsnProGlyGln 156

QY 545 -----TGACACCTGCAATGC-----TGCCACGACGAGGC 519

Db 157 AspAlaSerGluAspGlnThrAlaGlyProAspSerPheCysPheAspCysProAlaGly 176

QY 518 CCTGGAGGCTGGTACAGAGTGGCGGCGAGTCTCGGAGTTGAGTGTGGCAGCCTGAC 459

Db 177 Pro-----ProGlyProSerGlyAlaPro----- 184

QY 458 GGTGAGGCCACGCAAGTAACACAGGAGGTGACTGTACGCTCATAGTTCTGGTGCAGCC 399

Db 185 -----GlyGlnLysGlyProSerGlyAlaProGlyAlaProGlyGlnSerGlyGlyAla 202

QY 398 GCMGCTGTCAATTGAGGCTTGGCCACACATTCCAAGTTGACCGTGGCCCTGGGCGAGGTTT 339

Db 203 Ala-----LeuProGlyProProGlyPro----- 210

QY 338 CTGCCCCCAGTCGAGGAGGATTGAAGTCAGGCTCGTTGAAGGGGGCCCCAGGTAGTTCA 279

Db 211 -----AlaGlyProProGlyProAla 217

QY 278 GGTAGTCCACGTAAGTCGCGAGTTGATGCTCCAGGTAGCGGGTGAGGTGATAGTTT 219

Db 218 GlyGlnProGlySerAsnGlyAsnAlaGlyAlaProGlyAla----- 231

QY 218 TCTGGATGAGGGGCGGCGGCGGCGGCGGCTGATCTCTGTGCGATTAAAGAGCTGGCAGGGA 159

Db 232 -----ProGlyGlnValValAspVal-----Pro 239

QY 158 GGTGCCACGACCGTGCATAGGCAAGCTAAATCCCCACGAGTCCCTCTGTCGGAGGT 99

Db 240 GlyThrProGlyProAla-----GlyProProGlySerPro-----Gly 252

QY 98 CCATGGGCTGGGGCGCGGGCGGCGGCGGCTCTCTCCAGAGGCTGGCGGAGTG 39

Db 253 ProAlaGly---AlaProGlyGlnProGlyGlnAlaGlySerSerGlnProGlyGlyPro 271

QY 38 GGA-----GGCGAGCGGGCGGCTCCG 18

Db 272 GlyProGlnGlyAspAlaGlyAlaPro 280

RESULT 15

NME3 RAT

ID NME3 RAT STANDARD; PRT; 1237 AA.

AC Q00961;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Glutamate [NMDA] receptor subunit epsilon 3 precursor (N-methyl

GN GRIN2C.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

Result	Query	Score	Match	Length	DB	ID	Description
1	1224	81.7	225	11	Q9QZM3		Q9qzm3 mus muscul

SQ SEQUENCE 225 AA; 25261 MW; 68B1

ALIGNMENTS

Alignment Scores:
 Pred. No.: 1,2e-96 Length: 225
 Score: 1224.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.71% Indels: 0
 DB: 11 Gaps: 0

US-09-931-704-4 (1-819) x Q90ZM3 (1-225)

QY 95 ATGACCTCCGAGCGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTGTG 154
 Db 1 MetaspLeuArgAlaGlyAspSerTrpGlyMetLeuAlyCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGAGTGCAGCTCTTAATCGACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAenAthrGlyAspProGlyProSerIle 40

QY 215 CAGAAACCTATGACTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuAenAthrValLeuTrp 60

QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCCTGACTTCAATCCTCCTCGACTGGG 334
 Db 61 TyrLeuAenTyrLeuGlyProPheAenGluProAspPheAenProProArgLeuGly 80

QY 335 CGAGAACTCTGCCAGCGGCGGCTCACTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuAenAthrValLeuTrp 60

QY 395 CTGCGGCTGAGTGCAGCTCTTAATCGACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAenAthrGlyAspProGlyProSerIle 40

QY 455 AACCGCTCAGGCTGCCAGCTCACTCCGAGCTGAGTCCGAGCTGAGCTTCTGTACAGGCTC 514
 Db 121 AenArgGlnAlaAthrAlaGluLeuAenAthrGlyAspSerLeuAlyCysThrSerLeu 140

QY 515 CAGGCGCTGCTGGGCGGCGGCTCACTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuAenAthrValLeuTrp 60

QY 575 CCTCTCCAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 634
 Db 161 ProLeuProGlyThrGluProAlaThrValAspSerLeuAlyCysThrSerLeu 140

QY 635 AAGATGATGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACGGCTTAAGAGAGAGATGACGCTCCAGCTGCTGCTGCTGCTGCTGCTGCT 754
 Db 201 AspPheAenArgLeuLysLysLysMetGlnProProAlaAthrValLeuHisLeu 220

QY 755 GAGGACATGTTTC 769
 Db 221 GluAlaHisGlyPhe 225

RESULT 2
 Q9UBD9
 ID Q9UBD9 PRELIMINARY; PRT; 225 AA.
 AC Q9UBD9;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE Neurotrophin-1/B-cell stimulating factor-3 (Cardiotrophin-like
 DE cytokine) (Similar to cardiotrophin-like cytokine,
 DE neurotrophin-1/B-cell stimulating factor-3).
 GN CLOC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99432254; PubMed=10500198;
 RA Senaldi G., Varnum B.C., Samiento U., Starnes C., Lile J., Scully S.,
 RA Guo J., Elliott G., McNinch J., Shaklee C.L., Freeman D., Manu F.,
 RA Simonet W.S., Boone T., Chang M.-S.;
 RT "Novel neurotrophin-1/B cell-stimulating factor-3: A cytokine of the
 RT IL-6 family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11458-11463(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99382254; PubMed=10448081;
 RA Shi Y., Wang W., Yourey P.A., Gohari S., Zukauskas D., Zhang J.,
 RA Ruben S., Alderson R.P.;
 RT "Computational EST database analysis identifies a novel member of the
 RT neurotrophic cytokine family.";
 RL Biochem. Biophys. Res. Commun. 262:132-138(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX Hu X., Xu Y., Zhang B., Peng X., Yuan J., Qiang B.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF176912; AAF0092.1; -;
 DR EMBL; AF172854; AAD54284.1; -;
 DR EMBL; AF176911; AAF00991.1; -;
 DR EMBL; AY049779; AAL15436.1; -;
 DR EMBL; BC012939; AAH12939.1; -;
 SQ SEQUENCE 225 AA; 25176 MW; E2DD4B6280833B55 CRC64;

Alignment Scores:
 Pred. No.: 5.45e-94 Length: 225
 Score: 1193.00 Matches: 218
 Percent Similarity: 98.22% Conservative: 3
 Best Local Similarity: 96.89% Mismatches: 4
 Query Match: 79.64% Indels: 0
 DB: 4 Gaps: 0

US-09-931-704-4 (1-819) x Q9UBD9 (1-225)

QY 95 ATGACCTCCGAGCGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTGTG 154
 Db 1 MetaspLeuArgAlaGlyAspSerTrpGlyMetLeuAlyCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGAGTGCAGCTCTTAATCGACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAenAthrGlyAspProGlyProSerIle 40

QY 215 CAGAAACCTATGACTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuAenAthrValLeuTrp 60

QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCCTGACTTCAATCCTCCTCGACTGGG 334
 Db 61 TyrLeuAenTyrLeuGlyProPheAenGluProAspPheAenProProArgLeuGly 80

QY 335 CGAGAACTCTGCCAGCGGCGGCTCACTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuAenAthrValLeuTrp 60

QY 395 CTGCGGCTGAGTGCAGCTCTTAATCGACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAenAthrGlyAspProGlyProSerIle 40

QY 455 AACCGCTCAGGCTGCCAGCTCACTCCGAGCTGAGTCCGAGCTGAGCTTCTGTACAGGCTC 514
 Db 121 AenArgGlnAlaAthrAlaGluLeuAenAthrGlyAspSerLeuAlyCysThrSerLeu 140

QY 515 CAGGCGCTGCTGGGCGGCGGCTCACTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuAenAthrValLeuTrp 60

QY 575 CCTCTCCAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 634
 Db 161 ProLeuProGlyThrGluProAlaThrValAspSerLeuAlyCysThrSerLeu 140

QY 635 AAGATGATGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACGGCTTAAGAGAGAGATGACGCTCCAGCTGCTGCTGCTGCTGCTGCTGCT 754
 Db 201 AspPheAenArgLeuLysLysLysMetGlnProProAlaAthrValLeuHisLeu 220

QY 755 GAGGACATGTTTC 769
 Db 221 GluAlaHisGlyPhe 225

QY	805	GGGCTCGAGGTGTGGGGTTAAGGTTACAGAGGTACAGAAACCATGTGCCTCCCAAGTCGAG	746
Db	2	GlyProGlySerTyrGlyProSerGlyPro-GlyGlySerGlyAlaAlaAlaAlaAlaAla	21
QY	745	GGTGACTGAAGCTGCTGGAGGCTGCATCTTCTTAAAGCCGTTGAAGTCTCTGGAGTGA	686
Db	21	aAlaAlaSerGlyProGlyGly-GlnGlnGlyTyrGlyProGly---	40
QY	685	ACGCCATAGCCAGGCTCTCAGAGTCCCTTACAGCAGCCAGAGAATCATCCATCTTCTCGAGGAA	626
Db	40	erAlaAlaAlaAlaAlaAlaAlaGlyGlySerGlyProGlyGlyTyrGlyGlnGlyProS	60
QY	625	GTCACTGTGGGCA---GGGCCAGGGGGCCAGGCTGGCTCAGTCCCTGGCAGAGCGCTGGG	570
Db	60	erGlyTyrGlyProSerGlyProGlyAlaGlnGlnGlyTyrGlyProGlyGlyGlnGlyG	80
QY	569	GCAGTGGTAGCAAGCTCGCATGACACCTGCATCTGCCAGCAGAGCGCTCGGAGGC	510
Db	80	lySerGly-----AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlySer-GlyProGlyGly	97
QY	509	TGCTACAGAAAGTGGCCAGGCTACGTCGAGATTACGCTGTGGCAGCCTCAGCGTTGAGGC	450
Db	98	Tyr-----GlyProGlyAlaAlaGlyProGlyAsnTyrGlyProSerGlyProGly	114
QY	449	-----CAGCAAGTAACACAGGAGGTGAC-----TGT	423
Db	115	GlySerGlyAlaAlaAlaAlaSerAlaAlaAlaAlaSerGlyProGlyGlyGlnGlnGlyTyr	134
QY	422	ACGCCATAGTTCTGGGTCAGCCGAGCGCTGTCATTCAGGCTTCGCCACACATTCCTCAAGT	363
Db	135	GlyProGlyGlySerGlyAlaAlaAlaAlaAlaAlaAlaSerGlyGlyAla-----	150
QY	362	TGACCGTGGCCCTGGGAGAGTTCTGCCCCAGTCGAGAGGATTGAAGTCAGGCTCGT	303
Db	151	-----GlyProGlyArg	154
QY	302	TGAAAGGG---GCCCGAGTGTTCAGGTAGTCCGACGTAAGTGC-----GGA	255
Db	155	GlnGlnGlyTyrGlyProGlyGlySerGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAla	174
QY	254	GTGATCTCCAGGTAGC---GGTGAGGTATAGTTTTCTGATGGAGGGGCCAGGGC	198
Db	175	GlySerGlyProGlyGlyTyrGlyGlnGlyPro**GlyTyrGlyProGlyGlyGlnGly	194
QY	197	CTGATCTCTGTGCGATTGAAGCTGGCCTCAGCGAGGTGCCACACGCGTCGCATA	138
Db	195	-----GlySerGlyGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla	204
QY	137	GGCAAGCTAACATCCCCACG-----AGTCCCTGCTGGAGTCCATGGGGCTGGGGG	84
Db	205	AlaAlaSerGlyPro**GlyTyrGlyProGlyAlaAlaGlyProGlyAsnTyrGly	224
QY	83	CGGGCCGGCCGGC-----GCCGCTCCTCCAGCAGGCTGGCGAGTGGGA	36
Db	225	ProSerGlyProGlyGlySerGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaSerGlyProGly	244
QY	35	GGGCGA	30
Db	245	GlyGln	246
RESULT 5			
ID	Q9F2N5	PRELIMINARY;	PRT; 889 AA.
AC	Q9F2N5;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical protein SC03115.		
GN	SC03115 OR SCE41.24C.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		

Db 75 LeuProHisIleLysValAlaAsnLeuProThrAlaAlaMetAspTyrAspThrPheMet 94
Qy 380 AGCCTCAATGACAGCTGGCTGACCCAGAACTATGAGGCGTACAGTACCTCTGTGT 439
Db 95 LysGlnThrAspGluThrArgLeuAsnAsnLeuTyrPheTyrSerAlaIleValGlu 114
Qy 440 TACTTGGCT---GGCTCAACGCTCAG-----GCTGCCACAGCTGAATCCGA 484
Db 115 PheLeuLysGluAlaMetThrGluGlnGluAspLeuAsnProAlaGluLeuSerLeuLys 134
Qy 485 CQTAGCTGGCCCACTCTGTACACCTCCAGGCTCGTGGCGCAGCATTCGAGGTGC 544
Db 135 AlaLysPheGluGluAlaMetAlaAsnSerAsnThrLeuIleSerLysIleSerAspIle 154
Qy 545 ATGGCGACGCTGGTACCCACCTGCCCGCCAGCTCTGCCAGGAGTACGACGCTGGGCC 604
Db 155 MetThrGlnMetGlyMetSerValThrIleThrLeuPro-----LysProLeuValVal 172
Qy 605 CTTGCCCTGCCACAGTACTCTCTCCAGAGATGGATGACTTCTGGCTGTGAAGGAG 664
Db 173 ProPheGluGlySerAlaTyrPheArgLysLysLeuArgGlyGlyValValCysLysGlu 192
Qy 665 CTGCAGACCTGGCTATGGCTTCAGCCAGAGTCTCAACCGGCTTAAGAAGAAGATCCAG 724
Db 193 TyrLysGluArgValLeuLeuThrLysArgAspPheGluPheLeuAlaLysLysTyrGln 212
RESULT 11
Q9PUJO PRELIMINARY; PRT; 215 AA.
ID Q9PUJO 01-MAY-2000 (TrEMBLrel. 13, Created)
AC Q9PUJO 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Receptivity factor isoform 3 precursor.
GN PRF.
OS Plethodon jordani (Salamander).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;
OC Plethodon.
OX NCBI_TaxID=8336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420364; PubMed=10489368;
RA Rollmann S.M., Houck L.D., Feldhoff R.C.;
RT "proteinaceous pheromone affecting female receptivity in a terrestrial
RT salamander.";
RL Science 285:1907-1909 (1999).
DR EMBL; AF181482; AAF01027.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 1 23 POTENTIAL.
Alignment Scores: 1 23
Pred. No.: 0.000428 Length: 215
Score: 144.50 Matches: 44
Percent Similarity: 46.99% Conservative: 34
Best Local Similarity: 26.51% Mismatches: 73
Query Match: 9.65% Indels: 15
DB: 13 Gaps: 4
US-09-931-704-4 (1-819) x Q9PUJO (1-215)
Qy 260 AGCTTAGCTGGGACCTACTCACTGACCTGGGGCCCTTTCAACGAGCGCTGACTTCAAT 319
Db 55 SerLeuLeuProThrTyrLeuSerPheGlnGlyAlaProLeuSerAspProAspThrGln 74
Qy 320 CCTCTCGACTGGGGCAGAACTCTGCCAGGGCCACCGTCAACTGGAAGTGTGGCA 379
Db ,75 LeuProHisIleLysValValAlaAsnLeuProThrAlaAlaMetAspTyrAspThrPhe 94
Qy 380 AGCCTCAATGACAGCTGGCTGACCCAGAACTATGAGGCGTACAGTACCTCTGTGT 439
Db 95 ArgGlnThrAspGluThrArgLeuLysAsnLeuTyrPheTyrSerAlaIleValGlu 114

Qy 440 TACTTGGCT-----GGCTCAAC-----CGTCAGGCTGCCACA 472
Db 115 PheLeuLysGluAlaMetThrGluGlnGluAspLeuAsnProAlaGluLeuAlaLeuLys 134
Qy 473 GCTGAACCTCGACGTAGCTGGCCCACTTCTGTACACGCTCCAGGGCTGTGGGCGAGC 532
Db 135 AlaLysPheGluGluAlaMetAlaAsnSerAsnThr-----LeuIleSerLys 150
Qy 533 ATTGCAGGTCTCATGGCGACGCTGGTCTACCCACTGCCCCAGCTCTGCCAGGGAGTGTAG 592
Db 151 IleSerAspIleMetThrGlnMetGlyMetSerValThrIleThrLeuPro-----Lys 168
Qy 593 CCAGCTTGGGGCCCTGGCCCTCCACAGTACTTCTCTCCAGAGATGGATGACTTCTTGG 652
Db 169 ProLeuValValProPheLysGlySerAlaTyrPheSerLysLysLeuArgGlyGlyVal 188
Qy 653 CTGCTGAAGAGCTGACAGCTGGCTATGGCTTCCAGCAAGGACTTCAACCGGCTTAAG 712
Db 189 ValCysLysGluTyrLysGluArgValPheLeuThrLysArgAspPheMetLeuLeuAla 208
Qy 713 AAGAAGATGACGCTCCA 730
Db 209 GluLysTyrGlnGlyPro 214
RESULT 12
O41935 PRELIMINARY; PRT; 585 AA.
ID O41935 01-JAN-1998 (TrEMBLrel. 05, Created)
AC O41935 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JAN-1998 (TrEMBLrel. 14, Last annotation update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE Hypothetical 60.2 kDa protein.
GN GAMMAHV.M6.
OS murid herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=33708;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WUMS;
RX MEDLINE=97366649; PubMed=9223479;
RA Virgin H.W. IV, Latreille P., Wamsley P., Hallsworth K., Weck K.E.,
RA Dal Canto A.J., Speck S.H.;
RT "Complete sequence and genomic analysis of murine gammaherpesvirus
RT 68.";
RL J. Virol. 71:5894-5904 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=WUMS;
RA Latreille P., Wamsley P., Waterston R.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97553; AAB66392.1; -.
KW Hypothetical protein.
SQ SEQUENCE 585 AA; 60160 MW; 85610AA80C34827D CRC64;
Alignment Scores:
Pred. No.: 0.00067 Length: 585
Score: 143.00 Matches: 75
Percent Similarity: 35.74% Conservative: 14
Best Local Similarity: 30.12% Mismatches: 82
Query Match: 9.55% Indels: 78
DB: 12 Gaps: 12
US-09-931-704-4 (1-819) x O41935 (1-585)
Qy 19 GGAGCCGGCTGGCGCTCCCTCCACTCCG-----CCAGCTCTGGGAGAGGA 63
Db 334 GlyProAspProArgProProArgProProGluLeuGlyProGlySerProThrSer 353
Qy 64 GCGCGCCCGCCGGCGG----- 78
Db 354 ProAlaProSerArgAlaGlyAlaArgIleProAspLeuProGlyProLeuProSerTrp 373

QY	79	----	CCCGCGCCCCAGCCCATCGAGCTCCGAGCAGGCGACTCGTGGGGAGTGTAGCTTG	133
Db	374	GlyProAspProArgProPro		383
QY	136	CTATGACAGGTGCTGTGGACCTCCCTGCGAGTCCAGCTCTTAATGCGACAGAGA---	192	
Db	384	ProGluLeuGlyProGlySerProThrSerProAlaProSerArgAlaGlyAlaAArgile	403	
QY	193	---TCCAGGCGCTGGCCCTC---	240	
Db	404	ProAspLeuProGlyProLeuLeuProSerTrpGlyProAspProArgProProPro	423	
QY	241	CCT-----	255	
Db	424	ProGluLeuLeuGlyProGlySerProThrSerProAlaProSerArgAlaGlyAlaAArgile	443	
QY	256	CGGACGCTTAGCTGGGACCTACCTGAACCTA---CCTGGGGCCCC---	309	
Db	444	ProArgProSerArgProProProGluLeuGlyProGlySerProArgPheProArg-Pr	463	
QY	310	TGACTTCAATCCTCTCGACTG-----GGGGCGAGAAACTCTGCCAGGCGCACGGTCAA	363	
Db	463	oSerArgProProProGluLeuLeuGlyProGlySerProAspLeu-ProGlyProLeuProS	483	
QY	364	CTTGGAAGTGGCGAGCGCTCAATGACAGCTCGCGCTGACCCAGAACTATGAGCGCTA	423	
Db	483	erTrp-GlyProAspProProThrPheProAla-----ProSerArgAlaGlyAla	499	
QY	424	CAGTCACCTCTGTTACTTGGCTGGCCCTCAACCGCTCAGGCTGCCACAGCTGAACCTCG	483	
Db	500	ArgileProProValProProThrPheProAlaProSerArgAlaGlyAlaAArgilePro	519	
QY	484	ACGTAGCTGGCCACTTCTGTACACGCTCCAGGCGCTCTGGGCGAGCATTTGCAGGTGT	543	
Db	520	ArgProPro-----ArgProProProGluLeuGly-----	529	
QY	544	CATGGCAGCTTGGCTACCCACTGCCCGAGCTCTGCCAGGAGCTGACCCAGCTGGCG	603	
Db	530	-----ProGlySerPro-AspLeuProGlyProLeuProSerTrpG1	543	
QY	604	CCCTGGCGCTGCCACAGT 622		
Db	543	yProValProProProSer 549		
RESULT 13				
Q9PU19	ID	Q9PU19	PRELIMINARY; PRT; 215 AA.	
AC	AC	Q9PU19		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DE	01-MAY-2000	(TrEMBLrel. 13, Last annotation update)		
GN	Receptivity factor isoform 4 precursor.			
OS	Plethodon jordani (Salamander).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Caudata; Salamandroidea; Plethodontidae;			
OC	Plethodon.			
OX	NCBI_Taxid=8336;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99420364; PubMed=10489368;			
RA	Rollmann S.M., Houck L.D., Feldhoff R.C.;			
RT	"Proteinaceous pheromone affecting female receptivity in a terrestrial			
RT	salamander.";			
RL	Science 285:1907-1909(1999).			
DR	EMBL; AF181483; AAF01028.1; -.			
KW	Signal.			
FT	SIGNAL			
SQ	SEQUENCE 1 23 POTENTIAL.			
	SEQUENCE 215 AA; 24054 MW; AAN412135FFRA4E7F CRC64;			
Alignment Scores:				
Pred. No.:		0.000636	Length:	215

US-09-931-704-4 (1-819) x Q905K9 (1-608)	
QY 19	GGAGCGCGGCTCGCCTCCCACTCGCCAGCCTCTGGAGAGAGCGCGCGCGCGG 78
Db 68	GlyAlaGlyGlnArg--ProSerGlyProThrGlyGlyHisProAlaAlaProGlyAla 86
QY 79	CCCGCGCGCCAGCCG-----CATGGA 99
Db 87	ProGlyProArgSerProArgThrGluArgArgGlySerAlaGlnArgGlyHisPro 106
QY 100	CCTCGAGCAGGAGCTCGTGGGGATGTAGCTTCCTATGACGCGTGTGTGGCACCT 159
Db 107	ProProGlyAlaGly-----GlnArgProSerGlyProThrGlyGlyHisProAlaAla 124
QY 160	CCTCGAGTGCAGCTCTTAATGCGCAGAGATCAGGCGCTGGCCCTCCATCCAGAA 219
Db 125	ProGlyAlaProGlyProArgSerProArgThrGluArgArgGlySerAlaGlnArg 144
QY 220	AACCTATGACCTACCGCTACCTGGAGCA-----TCAACTCCGACGCTTAGC 267
Db 145	Gly-----HisProProGlyAlaGlyGlnArgProSerGlyPro-----Thr 159
QY 268	TGGGACCTACCTGAACCTACCTGGGGCGCCCTTTCAACGAGCCTGACTTCAATCCTCTCG 327
Db 160	GlyGlyHisProAlaAlaProGlyAlaProGlyProArgSer-----ProArgThr 176
QY 328	ACTGGGGGAGAACTCTGCCAG-----GGCAGCGTCAACTTGAAGTGGCGAAGCCT 384
Db 177	GluArgArgGlySerAlaGlnArgGlyHisProProGlyAlaGlyGlnArgPro 196
QY 385	CAATGACAGCTCGGCTGACCCAGAACTATGAGCGCTACAGTCACCT----- 432
Db 197	SerGlyProThrGlyGlyHisProAlaAlaProGlyAlaProGlyProArgSerProArg 216
QY 433	-----CCTGTTACTTGTGCTGGCTGCTCA 456
Db 217	ThrGluArgArgGlySerAlaGlnArgGlyHisProProGlyAlaGlyGlnArg 236
QY 457	CGGTCA-----GCCTGCCAGCTGAACCTCGAGTGTAGCTGGCCCACTTCTG--- 504
Db 237	ProSerGlyProThrGlyGlyHisProAlaAlaProGlyAlaProGlyProArgSerPro 256
QY 505	-----TAC 507
Db 257	ArgThrGluArgArgGlySerAlaGlnArgGlyHisProProGlyAlaGlyGln 276
QY 508	CAGCTCCAGCGCTGTGTGGCAGCATTCAGGTGTATGCGAGCTTGGCTACCCACT 567
Db 277	ArgProSerGlyProThrGlyGlyHis----- 285
QY 568	GCCCGACCTCTGCAGGAGTGGCCAGCTGGGCGCTGGCCCTGGCCCTGCCAGTGACTT 627
Db 286	ProAlaAla-ProGly-----AlaProGlyProArg-SerProArgT 299
QY 628	CCTCCAGAGATGATGATCTTCTGCTGCTGAAGAGCTGCAGA----- 671
Db 299	hrGluArgArg-----ArgGlySerAlaGlnArgGlyHisProp 312
QY 672	CTGGCTATGGCTTCAGCAGGACTTCAACCGG----- 707
Db 312	roProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyHisProAlaAlaProGlyA 332
QY 708	-----TTAAGAAGAGATGACGCTCCAG----- 731
Db 332	laProGlyProArgSerProArgThrGluArgArgGlySerAlaGlnArgGlyHisP 352
QY 732	-----CAGCTTCACTCACCTCGCTGAGGAGCAGCATGTTCTGACC 774
Db 352	roProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyHis----- 367
QY 775	TCTGACCTTAACCCACACCTCCAGGCCAGTCAGC 812
Db 368	ProAlaAlaProGlyAlaProGlyProArgSer 378

US-09-931-704-4 (1-819) x Q04397 (1-383)	
QY 16	GCCGGAGCGCGCTCGCTCCCTCCCACTCCGCCA----- 48
Db 55	SerGlyAlaAlaGlySerProGlnAlaProGlnThrArgValSerAlaSerArgAla 74
QY 49	-----GCCTCTGGGAGAGGA----- 63
Db 75	AspArgProArgAlaTrpArgLeuLeuGlyAlaSerArgGlyTrpPheCysProSer 94
QY 64	-----GCCCGCGCGCGCGCGCGCGC----- 108
Db 95	LeuCysProSerGluGluProGlyThrSerGlyThrProGluProLeuGlyProAlaSer 114
QY 109	AGG-----GGACTCGTGGGGGATGTAGCTTGGCTATGCAC----- 144
Db 115	ArgArgProGlyLeuArgSerProLeuSer---ProValysProLysGluCysLeu 133
QY 145	-----GGTGTGTGGCAGCTCCCTGCTGAGTGCAGCTCTTAATCGCACAGG 189

RESULT 15

Q04397

PRELIMINARY; PRT; 383 AA.

AC Q04397; 1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

DE Hypothetical BWRFL protein.

GN BWRFL OR BCRF2

OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus.

OX NCBI_TaxID=10377;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84270667; PubMed=6087149;

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,

RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,

RA Tuffnell P.S., Barrell B.G.;

RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";

RL Nature 310:207-211 (1984).

RN [2]

RP SEQUENCE OF 88-145 FROM N.A.

RX STRAIN=B95-8;

RX MEDLINE=88296424; PubMed=2841116;

RA Walls D., Gannon F.;

RT "The expression of novel antigens from the Epstein-Barr virus large

RT internal repeat";

RL EMBO J. 7:1191-1196 (1988).

DR EMBL; V01555; CAA24870.1; -

DR EMBL; V01555; CAA24871.1; -

DR EMBL; V01555; CAA24872.1; -

DR EMBL; V01555; CAA24873.1; -

DR EMBL; V01555; CAA24874.1; -

DR EMBL; V01555; CAA24869.1; -

DR EMBL; V01555; CAA24875.1; -

DR EMBL; X07816; CAA30675.1; -

DR EMBL; V01555; CAA24864.1; -

DR EMBL; V01555; CAA24865.1; -

DR EMBL; V01555; CAA24866.1; -

DR EMBL; V01555; CAA24867.1; -

DR EMBL; V01555; CAA24868.1; -

KW Hypothetical protein.

SQ SEQUENCE 383 AA; 39866 MW; 3FA433CE1EB04461 CRC64;

Alignment Scores:

Pred. No.: 0.000933

Score: 141.00

Percent Similarity: 31.96%

Best Local Similarity: 26.98%

Query Match: 9.41%

DB: 12

Length: 383

Matches: 92

Conservative: 17

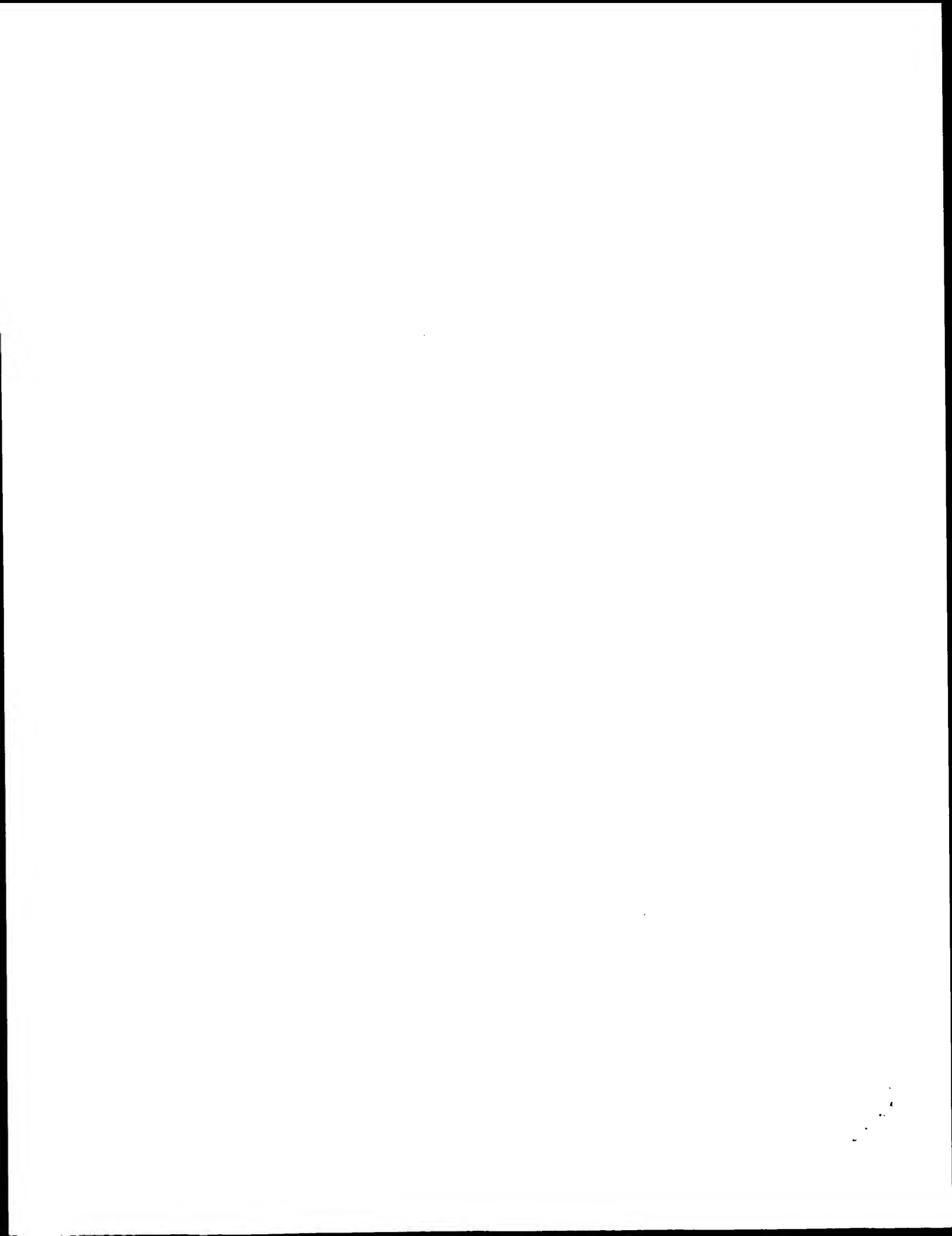
Mismatches: 97

Indels: 135

Gaps: 22

Db 134 ArgGlyAlaThrLeuGlyAlaGlnAlaProGluSerArgGlyGln-----GlyHisLeu 151
QY 190 AGATCCAGGCC-----|||||
Db 152 ArgValProProArgValProGlyGlnProGluGlyProArgGlnProGlyArgProGln 171
QY 220 AACCTATGACTACCGCTACCTGGAGCATCAACTCGCAGCTTAGCTGGGACCTACCT 279
Db 172 ArgProValProArgProGlyLeuGlnSerProGlyCysPro-----Pro 188
QY 280 GAA-----CTACCTGGGGCCCTTTCAACGAGCCTGACTTCAATCCTCCTCG 327
Db 189 GluGlyThrLeuGlyValProSerProPro-----LeuGlnAlaArgAla 203
QY 328 ACTGGGGGCAAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCCT--- 384
Db 204 SerProSerArgGlyAla-----SerLeuGlyProGlnValGlnProHis 219
QY 385 CAATGACAGGCTGGCGTGAACCCAGAACTATGAGGCGTACAGTCACTCTGTGTACTT 444
Db 220 ArgAspProSerGlyProAspProProThrGlyProSerLeuCysProProAlaProLeu 239
QY 445 GCCTGGCCTCAACCGTCAGGCTGCCACAGCTGAACCTCCGACGTAGCCTGGCCACTTCTG 504
Db 240 -----GlnProSer---LeuHis-----ProArgProGlnLeu 249
QY 505 TACAGCCTCCAGGCGCTGTGGGACGCAATTGCAGGTGTCTATGGCGACGCTTGCTACCC 564
Db 250 LeuAlaSerProGlyProProGlyGln-----Pro 259
QY 565 ACTGCCCCAGCCTCTGCCAGGGACTGAGCCAGCTGGGCCCTGG-----609
Db 260 GluGlyProArgGlnProGlyArgValAlaPheProLeuProTrpProLeuProAla 279
QY 610 -----CCCTGCCACAGTGACTTCTCTCCA-----GAA 636
Db 280 SerHisProSerProLeuSerLeuProProHisArgValHisGlnAlaGlyArgArgAsp 299
QY 637 GATGGATGACTTCTGGCTGTGAAGAGCTGCAGACCTGGCTATGGCGTTACGCCAAGGA 696
Db 300 ProGly-----GlyProValSerValProProAlaAlaGlnSer 313
QY 697 CTTCAACCGCTTAAGAAGATGACGCTCCAGCAGCTTCACTCACTGCACTTGA 756
Db 314 LeuProPro-----GlyLysGlyAlaSerPheSer-----323
QY 757 GGCACATGGTTTGTGACCTTACCTTAAACCCCACTCCAGGCCAGTCAGCTGTG 816
Db 324 -----ProProSerLeu-ArgProSerLeuLeuCy 333
QY 817 C 817
Db 333 s 333

Search completed: January 27, 2003, 16:20:19
Job time : 57.2562 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:43:12 ; Search time 34.4559 Seconds
(without alignments)
6334.602 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattataaagcttcgcgg.....agggccagtcagctgtgctt 819

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09931704/runat_27012003_154125_3585/app_query.fasta_1.7189
-DB=A Geneseq 101002 -QMT=fastan -SUFFIX=tag -MINMATCH=0.1 -LOPECL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704 @CNC_1_1_222 @runat_27012003_154125_3585 -NCPU=6 -ICPU=3
-NO XLPXY -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 101002.*

1: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/genesep-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/genesep-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/genesep-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/genesep-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/genesep-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/genesep-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/genesep-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/genesep-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/genesep-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/genesep-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/genesep-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/genesep-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/genesep-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/genesep-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/genesep-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/genesep-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/genesep-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/genesep-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/genesep-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/genesep-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/genesep-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/genesep-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1324	88.4	253	22	AAW25831
2	1224	81.7	225	19	AAW297116
3	1224	81.7	225	19	AAW56142
4	1224	81.7	225	21	AAW7814
5	1224	81.7	225	23	AAU78177
6	1193	79.6	225	19	AAW29715
7	1193	79.6	225	19	AAW56141
8	1193	79.6	225	20	AAW94466
9	1193	79.6	225	21	AAW87813
10	1193	79.6	225	22	AAW63543
11	1193	79.6	225	23	AAU78176
12	1192	79.6	321	22	ABH11896
13	1192	79.6	321	22	AAW79399
14	1185	79.1	260	22	AAW78415
15	1181	78.8	223	22	AAW80828
16	1167	77.9	215	21	AAW19587
17	1136	75.8	164	21	AAW19586
18	852	56.9	164	22	ABB40317
19	852	56.9	164	22	ABB24716
20	852	56.9	164	22	AAW61118
21	852	56.9	164	22	AAW73827
22	852	56.9	164	22	AAW20115
23	852	56.9	164	22	AAW34012
24	852	56.9	164	23	ABG43716
25	161	10.7	208	20	AAW09196
26	159	10.6	208	20	AAW09197
27	146	9.7	330	19	AAW57645
28	146	9.7	408	17	AAW07539
29	146	9.7	408	22	AAW72737
30	146	9.7	408	22	AAW64007
31	145.5	9.7	689	20	AAW08306
32	142	9.5	325	22	ABG21919
33	141.5	9.4	371	22	ABG05200
34	140.5	9.4	1064	17	AAW93254
35	140.5	9.4	1664	19	AAW37741
36	140.5	9.4	1065	14	AAW37741
37	139	9.3	684	18	AAW15287
38	137	9.1	574	22	AAW52322
39	137	9.1	574	22	AAW67370
40	137	9.1	772	13	AAW23873
41	137	9.1	772	18	AAW09643
42	137	9.1	1685	22	ABG04839
43	137	9.1	1693	22	ABG15619
44	136	9.1	682	16	AAW80330
45	136	9.1	682	19	AAW49717

ALIGNMENTS

RESULT 1

AAW25831

ID AAW25831 standard; Protein; 253 AA.

XX

AC AAW25831;

XX

DT 16-OCT-2001 (first entry)

XX Human protein sequence SEQ ID NO:1346.

DE

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;

XX anti-inflamatory; antirheumatic; antiarthritic; immunosuppressive;

XX antibacterial; endocrine; cardiant; central nervous system; virucide;

XX anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;

XX antiaggregant; haemostatic; vulnary; antitumor; osteopathic; sczema;

XX dermatological; antiallergic; antiaschmatic; antidiabetic; cytostatic;

XX neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;

XX immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;

XX antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.

OS WO200153455-A2.

PN 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

PF 23-DEC-1999; 99US-0471275.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-457603/49.

DR N-PSDB; AAH99772.

XX Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 20; Page 278; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to

CC AAM25963. The proteins can have activities based on the tissues and

CC cells they are expressed in, such as: antinflammatory; antirheumatic;

CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;

CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;

CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;

CC antiulcer; osteopathic; dermatological; antiallergic; antisthmatic;

CC antiadrenergic; cyostatic; neuroprotective; antidepressant; nootropic;

CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

CC encoding them can be used in gene therapy, antisense therapy and vaccine

CC production. The proteins and polynucleotides are useful for screening for

CC agonists or antagonists of a protein and for the treatment and diagnosis

CC of disorders associated with the activity of a protein e.g. inflammation,

CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal

CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,

CC osteoporosis, severe combined immunodeficiency, eczema, allergic

CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

CC Alzheimer's disease, Parkinson's disease, neurodegenerative and

CC neurological disorders.

XX SQ Sequence 253 AA;

Alignment Scores:

Pred. No.: 9.8e-114 Length: 253

Score: 1324.00 Matches: 245

Percent Similarity: 97.64% Conservative: 3

Best Local Similarity: 96.46% Mismatches: 5

Query Match: 88.38% Indels: 1

DB: 22 Gaps: 0

US-09-931-704-4 (1-819) x AAM25831 (1-253)

Qy 10 AGCTTGGCGGAGCGCGGCTCGCCCTCCACTCGCCAGCTCTGGGAGAGCGCGG 69

Db 1 SerPheAlaGlyAlaAlaAlaArgProSerThrProAlaSerGlyArgGlyAlaAla 20

Qy 70 CCCG 129

Db 21 ProGlyArgProGly-ProSerProMetAspLeuArgAlaGlyAspSerTrpGlyMetLe 40

QY 130 AGCTTGGCTATGACCGGTGTGTGGCCCTCCCTGCTGAGTGCAGCTCTTAATCGCACAGG 189
 DB 40 uAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAsnArgThrGl 60
 QY 190 AGATCCAGGCCCTGGCCCTCCATCCAGAAACCTATGACCTCACCCTGCTACCTGGAGCA 249
 DB 60 YAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTyrLeuGluHi 80
 QY 250 TCAACTCCGACGCTTAGCTGGGACCTACTGAACCTACCTGGGGCCCTTTTCAACGAGCC 309
 DB 80 sGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPheAsnGluPr 100
 QY 310 TGACTTCAATCCTCCTCGACTGGGGGAGAACTCTGCCAGGGCCACCGTCAACTTGA 369
 DB 100 oAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrValAspLeuGl 120
 QY 370 AGTGTGGGAGACCTCAATGACAGCGCTCGGCTGACCCAGGAGACTATGAGCGGTACAGTCA 429
 DB 120 uValTrpArgSerLeuAsnAspLysLeuArgLeuThrGlnAsnTyrGluAlaTyrSerHi 140
 QY 430 CCTCTGTGTACTTGTCTGCTGCTCAACCGCTCAGCTGCCAGCTGCCAGCTGAACTCCGAGCTAG 489
 DB 140 sLeuLeuCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGluLeuArgArgSe 160
 QY 490 CCTGGCCCATCTTCTACAGCTCCAGGCTCCAGGCTGCTGGCAGCATTCGAGTGTCTATGCC 549
 DB 160 rLeuAlaHisPheCysThrSerLeuGlnGlyLeuLeuGlySerIleAlaGlyValMetAl 180
 QY 550 GACGCTTGGCTACCCACTGCGCCAGCTCTGCGCAGGACTGAGCCAGCTGGGCGCCCTGG 609
 DB 180 aAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTrpThrProGl 200
 QY 610 CCCTGCCACACAGTGTCTTCTCCAGAGATGATGATCTTCTGGCTCTGAGGAGCTGCA 669
 DB 200 yProAlaHisSerAspPheLeuGlnLysMetAspPheTrpLeuLeuLysGluLeuGl 220
 QY 670 GACCTGGCTATGGCTTCAGCCAGGACTCAACCGGCTTAAGAAGAAGATGAGCGCTCC 729
 DB 220 nThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMetGlnProPr 240
 QY 730 AGCAGCTTCAGTCACCTGCTGAGGAGGACATGTTTC 769
 DB 240 oAlaAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 253

RESULT 2

AAW29716

ID AAW29716 standard; Protein; 225 AA.

AC AAW29716;

DT 09-NOV-1998 (first entry)

DE Mouse neurotrophic factor NNT-1.

XX NNT-1; neurotrophic factor; mouse; antiinflammatory; adjuvant;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome;
 KW peripheral neuropathy; dystrophy; neural retina degeneration;
 KW common variable immunodeficiency; CVID; selective IGA deficiency;
 KW hypogammaglobulinaemia; X-linked agammaglobulinaemia; antiseptic;
 therapy.

OS Mus sp.

XX Key Location/Qualifiers

FT Peptide 1..27

FT Protein /label= Sig_peptide

FT Protein /label= Mat_protein

XX WO9833922-A1.

XX 06-AUG-1998.

PD

XX 02-FEB-1998; 98WO-US02363.
 XX 30-JAN-1998; 98US-0016534.
 PR 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 XX Chang M, Elliot GS, Sarmiento U, Senaldi G;
 PI WPI; 1998-437475/37.
 DR N-PSDB; AAV47512.
 XX Newly isolated nucleic acid encoding human or murine neurotrophic
 PT factor NNT-1 - useful for treatment of neurological and
 PT immunological diseases or inflammation, also as vaccine adjuvant
 XX Claim 13; Fig 5; 120pp; English.
 XX This is the amino acid sequence of a murine neurotrophic factor,
 CC designated NNT-1, that is a growth factor for neurons and for B or
 CC T cells. It was deduced from isolated NNT-1 cDNA (see AAV47512).
 CC Human NNT-1 (see AAW29715) is also provided. Vectors and host cells
 CC for use in the production of human murine recombinant NNT-1
 CC polypeptides. These are used to treat: (i) neurological or
 CC immunological diseases, specifically Alzheimer's, Parkinson's
 CC or Huntington's diseases, amyotrophic lateral sclerosis,
 CC Charcot-Marie-Tooth syndrome, peripheral neuropathy, dystrophy and
 CC degeneration of the neural retina, or conditions characterised by T
 CC or B cell defects, e.g. common variable immunodeficiency (CVID),
 CC selective IgA deficiency, hypogammaglobulinaemia and X-linked
 CC agammaglobulinaemia (claimed), but many others disclosed; and (ii)
 CC antibody production following vaccination, and, since it inhibits
 CC tumour necrosis factor production, it may also be useful for
 CC treating sepsis. In addition, cells that have been engineered to
 CC express NNT-1 can be implanted, or nucleic acids are delivered in
 CC gene therapy vectors.
 XX SQ Sequence 225 AA;
 Alignment Scores:
 Pred. No.: 1-62e-104 Length: 225
 Score: 1224.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.71% Indels: 0
 DB: 19 Gaps: 0
 US-09-931-704-4 (1-819) x AAW29716 (1-225)
 QY 95 ATGGACCTCCGAGCGGACCTGCTGGGGAGTTAGCTTGCTATGCAGCGTGTCTGG 154
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 155 CACCTCCCTGCACTGCTTAAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
 QY 215 CAGAAACCTATGACCTACCCGCTACCTGGAGCACTCACTCCGACGCTTAGCTGGACC 274
 Db 41 GlnLysThrThrAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACGAGCCCTGACTCAATCTCTCGACTGGGG 334
 Db 61 TyrLeuAsnTyrLeuGlyProProAsnGluProAspPheAsnProProArgLeuGly 80
 QY 335 GCAGAAACCTGCCCCAGGCGGACGCTCACTTGGAGTGTGGGAGCCCTCAATGACAGG 394
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
 QY 395 CTGGCGGTGACCCAGAACTATGAGCGGTACAGTCACTCTGTGTACTTGGTGGCTC 454

Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 QY 455 AACCGTCAGGCTGCCACAGCTGAATCCAGCTAGCTGGCCCACTTCTGTACAGGCTC 514
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 515 CAGGCGCTGCTGGCAGCATTGTCAGGTGTATGCGGACGCTTGGCTACCCACTGCCCCAG 574
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 QY 575 CCTCTGCCAGGACTGAGCCAGCCTGGGCCCCCTGGCCCTGCCACAGTGTCTCTCCAG 634
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
 QY 635 AAGATGGATGACTTCTGCTGCTGAGGAGCTGCAGACCTGGCTATGGCGTTGAGCCAG 694
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 695 GACTTCAACCGCTTAAAGAAGATGCAAGCTCCAGCAGCTTCAGTCACTCCCTGCCTTG 754
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 QY 755 GAGGCACATGGTTTC 769
 Db 221 GluAlaHisGlyPhe 225
 RESULT 3
 AAW56142
 ID AAW56142 standard; Protein; 225 AA.
 AC AAW56142;
 XX 13-JUL-1998 (first entry)
 DT Amino acid sequence of murine neurotrophic factor NNT-1.
 DE Mouse; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
 DE treatment; neurological disease; degeneration; Parkinson's disease;
 DE amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
 XX Mus sp.
 XX Key Location/Qualifiers
 XX Peptide 1..27
 XX Protein /note= "signal peptide"
 XX 28..225
 XX /note= "mature peptide"
 XX US5741772-A.
 XX 21-APR-1998.
 XX 03-FEB-1997; 97US-0792019.
 XX 03-FEB-1997; 97US-0792019.
 XX (AMGE-) AMGEN INC.
 XX Chang M;
 XX WPI; 1998-260526/23.
 DR N-PSDB; AAV22654.
 XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
 PT useful for stimulating growth of motor and sympathetic neurons
 XX Claim 2; Fig 5; 41pp; English.
 XX The present sequence represents a murine neurotrophic factor, designated
 CC NNT-1, which is capable of stimulating growth of motor or sympathetic
 CC neurons. The NNT-1 protein is useful in the treatment of neurological
 CC diseases characterised by the degeneration and death of particular
 CC classes of neurons. These diseases specifically include Parkinson's

CC disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
 CC stroke and various degenerative disorders affecting vision.

XX
 SQ Sequence 225 AA;

Alignment Scores:

Pred. No.: 1,62e-104 Length: 225
 Score: 1224.00 Matches: 225
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.71% Indels: 0
 DB: 19 Gaps: 0

US-09-931-704-4 (1-819) x AAW56142 (1-225)

QY 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
 QY 215 CAGAAACCTATGACCTCACCCCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 275 TACCTGAACCTACCTGGGGCCCTTTCAAGAGCCCTGACTCAATCTCTCGACTGGG 334
 Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 335 CGAGAACTCTGCCAGGGCCAGGTCACTTGGAGTGTGGGAGCCCTCAATGACAGG 394
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
 QY 395 CTGCGGCTGACCCAGAACCTATGAGCGGTACAGTCACCTCTGTGTACTTGGCGTCC 454
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 QY 455 ACCGTGAGCTGCCAGCTGAACTCCGAGTAGCTGGCCCACTTCTGTACCAGCTC 514
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 515 CAGGGCTGTGCGGACGATTCAGGTGTGATGGCGACGCTTGGCTACCCACTGCCCCAG 574
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 QY 575 CTTCTGCCAGGACTGAGCCAGCTGGGCGCCCTGGCCCTGCCCAAGTACTTCTCCAG 634
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
 QY 635 AAGATGATGACTCTCTGGCTGCTGAAGGAGCTGCAGACCTGCTATGCGTTTACGCCAAG 694
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 QY 695 GACTTCAACCGGCTTAAGAGAAGATGACGCTCCAGCAGCTTTCAGTCCACCTGCACTTG 754
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 QY 755 GAGGCACATGGTTTC 769
 Db 221 GluAlaHisGlyPhe 225

RESULT 4

ID AAY87814
 AC AAY87814 standard; Protein; 225 AA.

XX

XX

XX

DT 24-AUG-2000 (first entry)

XX

DE Murine NNT-1 protein.

XX

KW NNT-1; neurotrophic factor; neurotropic; neuroprotective; treatment;
 KW anticonvulsant; antiparkinsonian; antidiabetic; ophthalmological;
 KW nervous system degeneration; Alzheimer's disease; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Charcot-Marie-Tooth syndrome; murine;
 KW Huntington's disease; peripheral neuropathy; neural retina degeneration;
 KW retinopathy; immune disorder; hematopoietic disorder.

XX Mus sp.

XX OS US6054294-A.

XX PN 25-APR-2000.

XX PD 12-DEC-1997; 97US-0988819.

XX PF 03-FEB-1997; 97US-0792019.

XX PR (AMGE-) AMGEN INC.

XX PA Chang M;

XX PI WPI; 2000-338492/29.

XX XX N-PSDB; AAA39483.

XX DR

XX PT

XX PT

XX PT

XX XX

XX PS

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX SQ

Sequence 225 AA;

Alignment Scores:

Pred. No.: 1,62e-104 Length: 225
 Score: 1224.00 Matches: 225
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.71% Indels: 0
 DB: 21 Gaps: 0

US-09-931-704-4 (1-819) x AAY87814 (1-225)

QY 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154

Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214

Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40

QY 215 CAGAAACCTATGACCTCACCCCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274

XX

Db 41. GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 275 TACCTGAACCTACCTGGGCCCCCTTTCAACAGAGCTGACTTCAATCCCTCCGACTGGG 334
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProPheArgLeuGly 80
 QY 335 GCAGAACTCTGCCAGGCCACGGTCACTTGGAGTGTGGCGAGCTCAATGACAGG 394
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTyrArgSerLeuAsnAspArg 100
 QY 395 CTGGGGCTGACCCAGAACTATGAGGCGTACACTCAGCTCCCTGTGTACTTGGTGGCTC 454
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 QY 455 AACCGTCAGGCTGCCACAGCTGAATCCGAGCTAGCTGGCCCACTTCTGTACACAGCTC 514
 Db 121 AsnArgGlnAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 515 CAGGCGCTGGGAGCAGATTGCGAGTGTCTATGGCGAGCTTGGTACCCACTGCCCCAG 574
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 QY 575 CCTCTGCAGGACTGAGCAGCTGGGCCCTCGCCCTGCGCTGAGCTGCTGCTGCTGCTG 634
 Db 161 ProLeuProGlyThrGluProAlaTyrAlaProGlyProAlaHisSerAspPheLeuGln 180
 QY 635 AAGATGATGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 694
 Db 181 LysMetAspAspPheTyrLeuLeuLysGluLeuGlnThrTyrPleuTyrPheSerAlaLys 200
 QY 695 GACTTCAACCGCTTATAGAGAAGATGAGCTCCAGCTCCAGAGCTTCACTGCTGCTGCTG 754
 Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 QY 755 GAGGCACATGCTTC 769
 Db 221 GluAlaHisGlyPhe 225

RESULT 5
 AAU78177
 ID AAU78177 standard; Protein; 225 AA.
 XX
 AC AAU78177;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Mouse novel neurotrophic factor NNT1.
 XX
 KW Mouse; NNT1; neurotrophic factor; IgE-related disease;
 KW Type 1 allergic disease; allergic rhinitis; eczema; dermatitis;
 KW pollinosis; asthma; immune disease; cancer; arteriosclerosis;
 KW vascular restenosis; rheumatoid arthritis; psoriatic arthritis;
 KW inflammatory arthritis; osteoarthritis; inflammatory joint disease;
 KW autoimmune disease; multiple sclerosis; lupus; diabetes; endometriosis;
 KW inflammatory bowel disease; transplant rejection; reproductive disorder;
 KW graft versus host disease; infertility; miscarriage; preterm labour.
 XX
 OS Mus sp.
 XX
 PN WO200215977-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 17-AUG-2001; 2001WO-US25906.
 XX
 PR 18-AUG-2000; 2000US-226436P.
 PR 16-AUG-2001; 2001US-0931704.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Senaldi G;
 XX
 DR WPI; 2002-280867/32.

N-PSDB; ABK11649.

Treating Immunoglobulin E-related disease, modulating IgE levels in a patient, preventing IgE-related disease and treating allergic diseases, involves administering NNT-1 inhibitor to a patient -

Claim 2; Fig 5; 63pp; English.

The invention relates to treating Immunoglobulin E (IgE)-related disease, modulating IgE levels in a patient, preventing an IgE-related disease, and treating allergic diseases, comprising administering a therapeutically effective amount of novel neurotrophic factor (NNT)-1 inhibitor to a patient. Also included are a method of diagnosing an IgE-related disease or susceptibility to an IgE-related disease, by determining the presence or amount of expression of an NNT1 polypeptide encoded by a NNT1 nucleotide sequence, its fragment or naturally occurring variant, and diagnosing an IgE-related disease or susceptibility of an IgE-related disease based on the presence or amount of expression of the polypeptide and a pharmaceutical composition for use in treating IgE-related disease, comprising the NNT1 inhibitor. The NNT1 inhibitor is useful for preventing and treating IgE-related disease, modulating IgE levels, and treating allergic diseases e.g. Type 1 allergic disease, allergic rhinitis, eczema, dermatitis, pollinosis, asthma, immune diseases and disorders, diseases involving abnormal cell proliferation including cancer, arteriosclerosis and vascular restenosis, diseases and conditions relating to dysfunction of immune system including rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, osteoarthritis, inflammatory joint disease, autoimmune disease, multiple sclerosis, lupus, diabetes, inflammatory bowel disease, transplant rejection, and graft versus host disease, and reproductive diseases and disorders including infertility, miscarriage, preterm labour and delivery, and endometriosis. The present sequence represents Mouse NNT1.

SQ Sequence 225 AA;

Alignment Scores:

Pred. No.: 1-62e-104 Length: 225
 Score: 1224.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.71% Indels: 0
 DB: 23 Gaps: 0

US-09-931-704-4 (1-819) x AAU78177 (1-225)

QY 95 ATGGACTCTGACGAGGGGACTCTGGGGATGTTAGCTTGCCTATGACGGTCTGTGG 154
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 155 CACCTCCCTGCGAGTCCAGCTCTTAATCCGACAGGAGATCCAGGCCCTGCCCTCCATC 214
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
 QY 215 CAGAAACCTATGACCTCACCCCTACCTGGAGCATCACTCCGAGCTTAGCTGGGACC 274
 Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAACAGAGCTGACTTCAATCCCTCCGACTGGG 334
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProPheArgLeuGly 80
 QY 335 GCAGAACTCTGCCAGGCCACGGTCACTTGGAGTGTGGCGAGCTCAATGACAGG 394
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTyrArgSerLeuAsnAspArg 100
 QY 395 CTGGGGCTGACCCAGAACTATGAGGCGTACAGTCACTCTCTGTGTACTTGGTGGCTC 454
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 QY 455 AACCGTCAGGCTGCCACAGCTGAATCCGAGCTAGCTGGCCCACTTCTGTACACAGCTC 514
 Db 121 AsnArgGlnAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140


```

Db 221 GlyAlaHisGlyPhe 225
RESULT 7
AAW56141
ID AAW56141 standard; Protein; 225 AA.
AC AAW56141;
XX
XX 13-JUL-1998 (first entry)
DE Amino acid sequence of human neurotrophic factor NNT-1.
KW Human; neurotrophic factor; NNT-1; growth; motor; sympathetic; neuron;
KW treatment; neurological disease; degeneration; Parkinson's disease;
KW amyotrophic lateral sclerosis; ALS; Alzheimer's disease; stroke.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /note= "signal peptide"
FT /note= "mature protein"
XX
XX US5741772-A.
PN
XX 21-APR-1998.
XX
XX 03-FEB-1997; 97US-0792019.
PR 03-FEB-1997; 97US-0792019.
XX (AMGE-) AMGEN INC.
XX Chang M;
XX WPI; 1998-260526/23.
XX N-PSDB; AAV22852.
XX Neurotrophic factor NNT-1 polypeptide and related nucleic acids -
XX useful for stimulating growth of motor and sympathetic neurons
XX Claim 1; Fig 3; 41pp; English.
XX
XX The present sequence represents a human neurotrophic factor, designated
XX NNT-1, which is capable of stimulating growth of motor or sympathetic
XX neurons. The NNT-1 protein is useful in the treatment of neurological
XX diseases characterised by the degeneration and death of particular
XX classes of neurons. These diseases specifically include Parkinson's
XX disease, amyotrophic lateral sclerosis (ALS), Alzheimer's disease,
XX stroke and various degenerative disorders affecting vision.
XX
SQ Sequence 225 AA;
Alignment Scores:
Pred. No.: 1,18e-101 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 79.64% Indels: 0
DB: 19 Gaps: 0
US-09-931-704-4 (1-819) x AAW56141 (1-225)
Qy 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCCATGACGGTGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
Qy 155 CACCTCCCTGCGAGTCCAGCTTTAATCCACAGGAGATCCAGGCCCTGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
QY 215 CAGAAAACCTATGACCTCACCGCTACCTGGAGCATCAACTCCCGAGCTTAGCTGGGACC 274
DB 41 GlnLysThrTrpAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAACGAGGCTGACTTCAATCTCTCTCGACTGGGG 334
DB 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
QY 335 GCAGAAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394
DB 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
QY 395 CTGCGGCTCACCCAGAACTATGAGGGGTACAGTCACCTCTGTTACTTTCGCTGGGCTC 454
DB 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
QY 455 AACCGTCCAGGTCGCCACAGCTCAACTCCGACGTAGCTGGCCCTCCACAGTCTTCTGTACCAGCTC 514
DB 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
QY 515 CAGGGCTCTGCTGGGAGCATTCGAGGTGTATGATGGCGAGCTTGGCTACCCACTGCCCCAG 574
DB 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
QY 575 CCTCTGCCAGGACTGAGCCAGCTGGGCCCTCTGGCCCTCCACAGTCTTCTGTACCAGCTC 634
DB 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
QY 635 AAGATGGATGACTTCTGGCTGTGAGAGAGCTGCAGACCTTGGCTATGGCGTTTCCAGCCAAG 694
DB 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
QY 695 GACTTCAACCGGCTTAAGAAGAGATGCGACCTCCAGCAGCTTCAGTACCCCTGCACCTTG 754
DB 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
QY 755 GAGGCACATGGTTTC 769
DB 221 GlyAlaHisGlyPhe 225
RESULT 8
AAW94466
ID AAW94466 standard; Protein; 225 AA.
XX
XX AAW94466;
XX
XX 22-APR-1999 (first entry)
DE Human cardiotrophin-like cytokine protein.
XX
XX Human; cardiotrophin-like cytokine; interleukin 6 cytokine family;
XX CLC; IL-6; diagnosis; detection; immune system-related disorder;
XX cancer; cardiac disorder; heart failure; hypertension; cancer;
XX autoimmune disorder; infection.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..27
XX Protein /label= signal
XX Domain 28..225
XX /label= Cardiotrophin-like_cytokine
XX /label= CD-I
XX /note= "conserved domain"
XX Domain 150..156
XX /label= CD-II
XX /note= "conserved domain"
XX Domain 194..198
XX /label= CD-III
XX /note= "conserved domain"
XX
XX WO9900415-A1.

```


CC hematopoietic disorders. (I) is effective in treating neurological
 CC conditions and promotes neuron regeneration. Neural functions are
 CC effectively restored in patients suffering from various neurological
 CC disorders. This sequence represents the human NNT-1 protein described in
 CC the method of the invention.

XX Sequence 225 AA;

Alignment Scores:
 Pred. No.: 1.18e-101 Length: 225
 Score: 1193.00 Matches: 218
 Percent Similarity: 98.22%
 Best Local Similarity: 96.89%
 Query Match: 79.64%
 DB: 21
 Gaps: 0

US-09-931-704-4 (1-819) x AAY87813 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTGTGG 154
 Db |||||
 1 MecAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 QY 155 CACCTCCCTCGAGCGCCAGCTCTTAATCGCACAGAGATCCAGGCCCTGCCCTCCATC 214
 Db |||||
 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
 QY 215 CAGAAACCTATGACCTACCCGCTA CCGTGAAGCATCACTCCGAGCTTAGCTGGGACC 274
 Db |||||
 41 GlnLysThrTrpAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 QY 275 TACCTGAACCTACCTGGGGCCCTTTCAACGAGCGCTGACTTCAATCTCTCGACTGGGG 334
 Db |||||
 61 TyrLeuAsnTrpLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
 QY 335 GCAGAACTCTGCCAGGGCCACGGTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394
 Db |||||
 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
 QY 395 CTGGGCTGACCCAGAACCTATGAGGCTACAGTCACTCTCTGTGTTACTTGGTGGCCTC 454
 Db |||||
 101 LeuArgLeuThrGlnAsnTrpGluAlaTrpSerHisLeuLeuGlyCysTyrLeuArgGlyLeu 120
 QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGAGCTAGCTGGCCCACTTCTGTACAGCTTC 514
 Db |||||
 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
 QY 515 CAGGGCTCTGGGAGCATTGTCAGGTGTCTATGCGCAGCTTGCTACCCACTGCCCCAG 574
 Db |||||
 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
 QY 575 CCTCTGCCAGGACTGAGCAGCTGGCCCTGGCCCTGCCACAGTACTTCTCTCCAG 634
 Db |||||
 161 ProLeuProGlyThrGluProTrpThrTrpProGlyProAlaHisSerAspPheLeuGln 180
 QY 635 AAGATGATGACTTCTGGTCTGAAGGAGCTGCAGACCTGGCTATGGCTTACGCCAAG 694
 Db |||||
 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnTrpLeuTrpArgSerAlaLys 200
 QY 695 GACTTCAACCGCTTAAGAGAGATGACGCTCCAGCAGCTTCAAGTCACTGCACTTG 754
 Db |||||
 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
 QY 755 GAGGCACATGGTTTC 769
 Db |||||
 221 GlyAlaHisGlyPhe 225

RESULT 10
 ID AAG63543
 AC AAG63543 standard; Protein; 225 AA.
 XX AAG63543;
 AC AAG63543;

DT 15-OCT-2001 (first entry)

XX

Amino acid sequence of a human NNT-1 protein.

NNT-1; CLF-1; sCNTFRalpha; nervous system; neuron; nervous system;
 neuro-muscular function; tumour; immune system; haematopoietic system;
 reproductive system; liver; skeletal muscle; neurodegenerative disease;
 amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
 muscular mass; paralysis; cancer; obesity; fertility; endometriosis;
 blastocyst implantation; thrombosis; retinal disease;
 retinal pigmentosis.

Homo sapiens.

WO200155172-A2.

02-AUG-2001.

26-JAN-2001; 2001WO-FR00253.

27-JAN-2000; 2000FR-0001035.

12-OCT-2000; 2000FR-0013089.

(FABR) FABRE MEDICAMENT SA PIERRE.

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

Elson G, Gauchat J, Plun-Pavreau H, Chevalier S, Gascan H;

WPI; 2001-488773/53.

N-PSDB; AAH74484.

A complex comprising a NNT-1 protein and a CLF-1 and/or sCNTFRalpha
 protein useful to treat neurodegenerative disease including Parkinson's
 and Huntington's, obesity and cancer.

Claim 2; Page 58; 67pp; French.

The present sequence represents a human NNT-1 protein. The specification
 describes a complex comprising a NNT-1 protein and a CLF-1 and/or
 sCNTFRalpha protein. The NNT-1/CLF-1 complex is used to modulate
 activity of the sCNTFRalpha/gp130/LiFRbeta receptor complex, or to
 induce phosphorylation of the tyrosine of gp130 and LiFRbeta,
 particularly where cells expressing the receptor complex are in the
 central or peripheral nervous system, in neurons implicated in
 neuro-muscular function or in skeletal muscle. The complex or
 antibodies are also used to decrease the survival, growth or
 proliferation of tumour cells or to facilitate the proliferation and/or
 inhibit differentiation of cells stocks. The complex is also used to
 modulate activity of the gp130/LiFRbeta receptor or cells expressing
 that receptor, particularly those cells implicated in the immune
 muscle. Molecules of the invention may be used to prevent or treat
 neurodegenerative diseases including amyotrophic lateral sclerosis,
 Parkinson's and Huntington's disease, to repair or regenerate nervous
 or muscular tissue or to maintain muscular mass in paralysis patients.
 They may also be used to treat cancer, obesity and associated diseases,
 and to improve fertility, particularly to avoid endometriosis and/or
 assist blastocyst implantation, thrombosis, or retinal disease,
 particular retinal pigmentosis.

Sequence 225 AA;

Alignment Scores:

Pred. No.: 1.18e-101 Length: 225
 Score: 1193.00 Matches: 218
 Percent Similarity: 98.22%
 Best Local Similarity: 96.89%
 Query Match: 79.64%
 DB: 22
 Gaps: 0

US-09-931-704-4 (1-819) x AAG63543 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTGTGG 154
 |||||

QY 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTCCAGGCGCCACGGT 360
Db 165 eAsnGluProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrVa 185
QY 361 CAACCTGGAAGTGTGGCGAGCTCAATGACAGGCTGGCGTGACCCAGAACATATGAGGC 420
Db 185 lAspLeuGluValTrpArgSerLeuAsnAspLysLeuA-gLeuThrGlnAsnTyrGluAl 205
QY 421 GTACAGTCACTTCTGTGTACTTGGTGGCTTCAACCCGTCAGGCTGCCACAGCTGAAC 480
Db 205 aTyrSerHisLeuLeuGlyCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGlu 225
QY 481 CCGAGTACCTGGCGCCACTTGTACCGAGCTCCAGGCTGCTGGCGAGCATTCGAGG 540
Db 225 uArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaG 245
QY 541 TGTCTATGGGAGCTGGCTTACCTACCTGCCCCAGGCTCTCCAGGAGTCCAGGCGCTG 600
Db 245 yValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTr 265
QY 601 GGGCCCTGGCCCTGCCACAGTGAATCTTCCAGAGATGATGACTTCTGGCTGTGAA 660
Db 265 pThrProGlyProAlaHisSerAspPheLeuGlnLysMetAspAspPheTrpLeuLeu 285
QY 661 GGAGCTGCAGACTGGCTATGGCTTCCAGCAAGGACTTCAACCGCTTAAAGAGAGAT 720
Db 285 sGluLeuGlnThrTrpLeuTrpArgSerAlaLysAspPheAsnArgLeuLysLysMe 305
QY 721 CGAGCTCCAGCAGCTTCAGTCACTCCCTGACCTGGAGGCACATGTTTC 769
Db 305 tGlnProProAlaAlaValThrLeuHisLeuGlyAlaHisGlyPhe 321

RESULT 13
ID AAM79399 standard; Protein; 321 AA.
XX AC AAM79399;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein SEQ ID NO 3045.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52532.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 20; Page 237; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX SQ Sequence 321 AA;
Alignment Scores: 1.61e-101 Length: 321
Pred. No.: 1192.00 Matches: 226
Score: 89.49% Conservative: 4
Percent Similarity: 87.94% Mismatches: 14
Best Local Similarity: 79.57% Indels: 13
Query Match: 22 Gaps: 3
DB:
US-09-931-704-4 (1-819) x AAM79399 (1-321)
QY 37 CCCACTCCGCGAGCTCTGGGAGA---GGAGCCGGCGCC---GGCGCGCCCGCCCCCAG 90
Db 66 ProHisProProSerProArgTrpGlyGlnThrProGluGlyLeuProAlaAlaSer 85
QY 91 CCCCATGGACCT-----CCGAGCAGGAGCTCTG 120
Db 86 ProCysGlyProGlyProArgSerCysPheSerSerIleLeuProThr-GlyAspSerTr 105
QY 121 GGGGATGTAGCTTGCCTATGCACGCTGTGGGACCTCCCTGCAGTGCAGCTCTTAA 180
Db 105 pGlyMetLeuAlaCysLeuCysThrValLeuTrpHisLeuProAlaValProAlaLeuAs 125
QY 181 TCGCACAGAGATCCAGGCGCTTCCATCCAGAAACCTATGACCTACCCGCTA 240
Db 125 nArgThrGlyAspProGlyProGlyProSerIleGlnLysThrTyrAspLeuThrArgTy 145
QY 241 CCTGAGCATCAATCCCGAGCTTAGCTGGGACCTACCTGAACTACTGGGGCCCCCTTT 300
Db 145 rLeuGluHisGlnLeuArgSerLeuAlaGlyThrTyrLeuAsnTyrLeuGlyProProPh 165
QY 301 CAACGAGCTGACTTCAATCTCTCGACTGGGGGAGAACTCTCCAGGCGCCACGGT 360
Db 165 eAsnGluProAspPheAsnProProArgLeuGlyAlaGluThrLeuProArgAlaThrVa 185
QY 361 CAACCTGGAAGTGTGGCGAGCTCAATGACAGGCTGGCGTGACCCAGAACATATGAGGC 420
Db 185 lAspLeuGluValTrpArgSerLeuAsnAspLysLeuA-gLeuThrGlnAsnTyrGluAl 205
QY 421 GTACAGTCACTTCTGTGTACTTGGTGGCTTCAACCCGTCAGGCTGCCACAGCTGAAC 480
Db 205 aTyrSerHisLeuLeuGlyCysTyrLeuArgGlyLeuAsnArgGlnAlaAlaThrAlaGlu 225
QY 481 CCGAGTACCTGGCGCCACTTGTACCGAGCTCCAGGCTGCTGGCGAGCATTCGAGG 540
Db 225 uArgSerLeuAlaHisPheCysThrSerLeuGlnGlyLeuGlySerIleAlaG 245
QY 541 TGTCTATGGGAGCTGGCTTACCTACCTGCCCCAGGCTCTCCAGGAGTCCAGGCGCTG 600
Db 245 yValMetAlaAlaLeuGlyTyrProLeuProGlnProLeuProGlyThrGluProThrTr 265

XX Human; biologically active complex; haemopoietin receptor; NR6;
 KW cardiotrophin-like cytokine; CLC; therapy; prophylaxis; proliferation;
 KW differentiation; cell survival; neurotrophic activity.
 XX
 OS Homo sapiens.

Key Location/Qualifiers
 FH Peptide 1..27
 FT /label= Signal_peptide
 FT Protein 28..223
 FT /label= Human mature CLC protein
 FT /note= "Cardiotrophin-like cytokine"
 FT
 XX

WO200127157-A1.

19-APR-2001.

06-OCT-2000; 2000WO-AU01216.

08-OCT-1999; 99AU-0003327.

12-MAY-2000; 2000AU-0007489.

(AMRA-) AMRAD OPERATIONS PTY LTD.

Nash A, Jachno KM, Fabri LJ, Reid K, Bartlett PF, Hilton DJ;

PI Nakata Y, Hasegawa M;

DR WPI; 2001-281978/29.

DR N-PSDB; AAD04201.

XX New biologically active complex comprising NR6 and

PT cardiotrophin-like-cytokine, for facilitating proliferation,

PT differentiation and/or survival of a cell -

XX

PS Claim 32; Page 114-115; 123pp; English.

XX The present invention relates to a biologically active complex comprising
 CC a haemopoietin receptor, NR6 and cardiotrophin-like cytokine (CLC).

CC The complex is useful in the manufacture of a medicament for the

CC treatment and/or prophylaxis of a subject, as it is involved in

CC facilitating proliferation, differentiation and/or survival of a cell.

CC The complex or its components have neurotrophic activity. The present

CC sequence is human cardiotrophin-like cytokine (CLC) protein.

XX Sequence 223 AA;

Alignment Scores:

Pred. No.:	1.5e-100	Length:	223
Score:	1181.00	Matches:	216
Percent Similarity:	98.21%	Conservative:	3
Best Local Similarity:	96.86%	Mismatches:	4
Query Match:	78.84%	Indels:	0
DB:	22	Gaps:	0

US-09-931-704-4 (1-819) x AAE00828 (1-223)

QY 95 ATGACCTCCGAGCAGCGGACTCGTGGGGATGTTAGCTTGCTATGACGGTGTGGG 154

Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyThrValLeuTrp 20

QY 155 CACCTCCCTGACGAGCCAGCTCTTAAATGCACAGAGATCCAGGCCCTGGCCCTCCATC 214

Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40

QY 215 CAGAAACCTATGACCTACCGGCTACCTGGAGCATCACTCCGAGCTTAGCTGGGACC 274

Db 41 GlnIysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAACCTACCTGGGCCCTTTCAACGAGCCTGACTTCAATCCTCTCGACTGGGG 334

Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAAACTCTGCCAGGGCCACCGTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394

Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100

QY 395 CTGGCGCTGACCCAGAACTATGAGCGGTACAGTCACCTCCTGTTACTTGGTGGCCCTC 454

Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 AACCGTCCAGGCTGCCACAGCTGAACTCCGAGCTGAGCCTGGCCCACTTCTGTACCAAGCCTC 514

Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGCCCTGTGGCAGCATTGCGGTGTCATGGCGAGCGCTTGGCTACCCACTGCCCCAG 574

Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160

QY 575 CCTCTGCCAGGACTGAGCCAGCCTGGGCCCTGGCCCTGCACAGTCACTTCTCTCCAG 634

Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGGATGACTTCTGCTGCTGAAAGGAGCTGCAGACCTGGCTATGGCGTTTCAGCCAG 694

Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAAGAAGATGCAGCCTCCAGCAGCTTCAGTCACCTGCACCTTG 754

Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220

QY 755 GAGGCACAT 763

Db 221 GlyAlaHis 223

Search completed: January 27, 2003, 15:59:53

Job time : 38.4559 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 16:20:23 ; Search time 5.86484 Seconds
(without alignments)
5635.701 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattattaaagcttcgcgg.....aggccagtcagctgtgctt 819

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 244452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USFTQ spool/US09931704/runat_27012003_154129_3760/app.query.fasta_1.7189
-DB=Published_applications_AA -QFW=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09931704 @CGN 1.1 51 @runat_27012003_154129_3760
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	81.7	225	10	US-09-931-704-5
2	1193	79.6	225	10	US-09-931-704-2
3	852	56.9	164	10	US-09-864-761-40014
4	134	9.2	529	10	US-09-861-597-2

Sequence	2,	Appli
Sequence 2,	Appli	
Sequence 50,	Appli	
Sequence 47,	Appli	
Sequence 56,	Appli	
Sequence 4,	Appli	
Sequence 2,	Appli	
Sequence 17,	Appli	
Sequence 217,	Appli	
Sequence 217,	Appli	
Sequence 1,	Appli	
Sequence 70,	Appli	
Sequence 2,	Appli	
Sequence 8,	Appli	
Sequence 10,	Appli	
Sequence 2,	Appli	
Sequence 5,	Appli	
Sequence 19,	Appli	
Sequence 720,	Appli	
Sequence 138,	Appli	
Sequence 14,	Appli	
Sequence 979,	Appli	
Sequence 14,	Appli	
Sequence 24,	Appli	
Sequence 1002,	Appli	
Sequence 3,	Appli	
Sequence 4,	Appli	
Sequence 3,	Appli	
Sequence 2,	Appli	
Sequence 9,	Appli	
Sequence 2,	Appli	
Sequence 216,	Appli	
Sequence 216,	Appli	
Sequence 34988,	Appli	
Sequence 36844,	Appli	
Sequence 3,	Appli	
Sequence 164,	Appli	
Sequence 254,	Appli	
Sequence 1422,	Appli	
Sequence 2,	Appli	

ALIGNMENTS

RESULT 1
US-09-931-704-5
; Sequence 5, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating IgE-Related Disease Using
; TITLE OF INVENTION: Inhibitors
; FILE OF INVENTION: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 225
; TYPE: PPT
; ORGANISM: Murine
US-09-931-704-5

Align. No.:	4,54e-90	Length:	225
Score:	1224.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	81.71%	Indels:	0
DB:	10	Gaps:	0

US-09-931-704-4 (1-819) x US-09-931-704-5 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGCACTGACCTTAAATCGCACAGGAGATCCAGGCCCTGCGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACTCCCGCTACCTGGAGCATCAACTCCGCGAGCTTAGCTGGGACC 274
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAATCTGCTGGGCGCCCTTTCAACGAGCCTGACTTCAATCCTCCCTCAGCTGGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAAACTCTGCCAGGCCCGCTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspArg 100

QY 395 CTGGGCTGACCCAGCACTATGAGCGGTACAGTCACCTCTCTGTGTACTTGGCTGGCCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 RACCGCTCAGGCTGCCACAGCTGAACCTCCGAGTAGCTGGCCCACTTCTGTACCAAGCCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGGCTCTGCGGACGATTCAGGTGTCTATGCGAGCTGAGCTGGCCCACTTCTGTACCAAGCCTC 574
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CCTGTCCAGGAGTACAGCAGCTGGGCGCCCTGGCCCTGCGCCAGTACTTCTCCAG 634
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AGATGGAGTACTTCTGGCTGCTGAAGGAGCTGCAGAGCTGAGCTGGCTGATGGCGTTCAGCCCAAG 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAGAGATGCGACCTCCAGCAGCTTCAGTCACCCCTGCACTTG 754
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220

QY 755 GAGGCACATGGTTTC 769
Db 221 GluAlaHisGlyPhe 225

RESULT 2
US-09-931-704-2
; Sequence 2, Application US/09931704
; Patent No. US20020041873A1
; GENERAL INFORMATION:
; APPLICANT: Senaldi, Giorgio
; TITLE OF INVENTION: Methods and Compositions for Treating Igb-Related Disease Using N
; TITLE OF INVENTION: Inhibitors
; FILE REFERENCE: A-695
; CURRENT APPLICATION NUMBER: US/09/931,704
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/226,436
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-931-704-2
Alignment Scores:

Pred. No.: 1.33e-87 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 79.64% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-4 (1-819) x US-09-931-704-2 (1-225)

QY 95 ATGGACCTCCGAGCAGGGGACTCGTGGGGATGTTAGCTTGCCTATGACGGTGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20

QY 155 CACCTCCCTGCACTGACCTTAAATCGCACAGGAGATCCAGGCCCTGCGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACTCCCGCTACCTGGAGCATCAACTCCGCGAGCTTAGCTGGGACC 274
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAATCTGCTGGGCGCCCTTTCAACGAGCCTGACTTCAATCCTCCCTCAGCTGGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAAACTCTGCCAGGCCCGCTCAACTTGAAGTGTGGCGAAGCCTCAATGACAGG 394
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100

QY 395 CTGGGCTGACCCAGCACTATGAGCGGTACAGTCACCTCTCTGTGTACTTGGCTGGCCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 AACCGTCCAGGCTGCCACAGCTGAACCTCCGAGTAGCTGGCCCACTTCTGTACCAAGCCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGGCTCTGCGGACGATTCAGGTGTCTATGCGAGCTGAGCTGGCCCACTTCTGTACCAAGCCTC 574
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CCTGTCCAGGAGTACAGCAGCTGGGCGCCCTGGCCCTGCGCCAGTACTTCTCCAG 634
Db 161 ProLeuProGlyThrGluProTrpTrpThrProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGGATGACTTCTGGCTGCTGAAGGAGCTGCAGAGCTGAGCTGGCTATGGCGTTCAGCCCAAG 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAGAGATGCGACCTCCAGCAGCTTCAGTCACCCCTGCACTTG 754
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220

QY 755 GAGGCACATGGTTTC 769
Db 221 GlyAlaHisGlyPhe 225

RESULT 3
US-09-864-761-40014
; Sequence 40014, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

```

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40014
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005849.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EST HUMAN HIT: AI752561.1, EVALUATE 3.00e-66
; OTHER INFORMATION: SWISSPROT HIT: Q63086, EVALUATE 8.00e-03
US-09-864-761-40014

Alignment Scores:
Pred. No.: 1.68e-60 Length: 164
Score: 852.00 Matches: 156
Percent Similarity: 97.56% Conservative: 4
Best Local Similarity: 95.12% Mismatches: 4
Query Match: 56.88% Indels: 0
DB: 10 Gaps: 0

US-09-931-704-4 (1-819) x US-09-864-761-40014 (1-164)
QY 278 CTGAACCTACCTGGGCCCCCTTCAACGAGCTGACTTCAATCTCTCGACTGGGGCA 337
Db 1 LeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGlyAla 20
QY 338 GAACTCTGCCAGGGCCACGCTCAACTTGGAGTGTGGCGAAGCCTCAATGACAGGCTG 397
Db 21 LysThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLysLeu 40
QY 398 CGGCTGACCCAGAACTATGAGCGGTACAGTACCTCTGTGTACTTGGTGGCCTCAAC 457

; PRIOR APPLICATION NUMBER: US 09861597
; Sequence 2, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-2

Alignment Scores:
Pred. No.: 0.00266 Length: 529
Score: 134.00 Matches: 74
Percent Similarity: 38.08% Conservative: 17
Best Local Similarity: 30.96% Mismatches: 97
Query Match: 9.17% Indels: 51
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-861-597-2 (1-529)
QY 671 TCTGCAGCTCTTCCAGCAGCCAGAGTCCATCTTCTGGAGGAAGTCAC----- 621
Db 73 SerAlaAlaAlaAlaAlaAlaGlySerGlyGlnGlnGlyProGlyGlyTyrGlyProArg 92
QY 620 -----TGTCGGCAGGCGC---CAGGGCCCGCAGGCTGGCTCAGTCCCTG 582
Db 93 GlnGlnGlyProGlyGlyTyrGlyGlnGlnGlyGlnGlnGlyProSerGlyProGlySerAla 112
QY 581 GCAGAGGCTGGGGCAGTGGGTAGCCAAAGCGTCGCCATGACACCTCAATGCTGCCAGCA 522
```

```
Db 113 AlaAlaIaSerAlaAlaSerAlaGluSer-----Gly 124
Qy 521 GGCCCTGGAGGCTGTACAGAGTGGCCAGAGTACGTGCGAGTTTCAGCTGTGGCAGCCT 462
Db 125 GlyProGlyGlyTyr-----GlyProGlyGlnGlnGlyProGlyGlyTyrGlyPro 141
Qy 461 GACGGTTGAGGCCACGCAAGTAACACAGGAGGTGACTGTACGCTTCATGTTCTGGGTCA 402
Db 142 GlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerGlyProGlySer 161
Qy 401 GCCCAGGCTGTATAGGCTTCCACACTTCCCAAGTTGACCGTGGCCCTGGCAGAG 342
Db 162 AlaAlaAlaAlaAla-----AlaAlaAlaSerGlyProGlyGlnGlnGlyProGly 179
Qy 341 TTTCTGCCCCAGTCA-----GGAGGATTGAAGTCAGGCTCGTTG---AAAGGGG 294
Db 179 TyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyProSerG 199
Qy 293 GCCCAGGTTAGTTTCAGGTAGTCCAGTAAAGCTGC-----GGAGTTGATGCTCCA 243
Db 199 lyProGlySerAlaAlaAlaAlaAlaAlaAlaSerGlyProGlyGlnGlnGlyPro 218
Qy 242 GGTAGC---GGGTAGGTCATAGTTTCTGGATGAGGGGCCAGGCGCTGGATCTCCTG 186
Db 219 GlyGlyTyrGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGly----- 234
Qy 185 TCGGATTAGAGCTGGCACTCGAGGAGGTGCCACAGCACCGTGCATAGCAAGCTAACCA 126
Db 235 -----GlnGlnGlyLeuSerGlyProGlySerAlaAlaAlaAla 247
Qy 125 TCCCCACAGTCCCTGCTCGAGGTCCATGGGGCTGGGGCGCGCGCGCGCGCG 66
Db 248 AlaAlaAlaGlyProGlyGlnGlnGlyProGlyGlyTyrGlyProGlyGlnGlnGlyPro 267
Qy 65 GTCCTCTCCAGAGCTGGCGAGTGGAGGCGGAGCGCGCTCGCGG 15
Db 268 Ser-----GlyProGlySerAlaAlaAlaAlaAlaAla 278

RESULT 5
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US2002004813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehlend
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: Motility
; CURRENT APPLICATION NUMBER: US/09/823,240
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-823-240-2

Alignment Scores:
Pred. No.: 0.00783 Length: 802
Score: 128.50 Matches: 73
Percent Similarity: 39.82% Conservative: 17
Best Local Similarity: 32.30% Mismatches: 82
Query Match: 8.58% Indels: 54
DB: 10 Gaps: 15

US-09-931-704-4 (1-819) x US-09-823-240-2 (1-802)
Qy 34 CCTCCACTCCGCCA-----GCCTCTGGGAGAGAGCGCGCGCGCGC---CGGCCC 81
```

```
Db 381 ProProSerProProIleMetIleSerSerProProGlyLysAlaThrGlyProArgPro 400
Qy 82 -----GGCCCCCAGCCCCCATGGACCTCCGAGC----- 108
Db 401 ValLeuProValCysValSerSerProValProGlnMetProProSerProThrAlaPro 420
Qy 109 AGGGGACTCGTGGGGATGTAGCTGCCTATG-----CAGGTCTCTGTG 153
Db 421 AsnGlySerLeuAaspSerValThrTyrProValSerProProProThrSerGlyProAla 440
Qy 154 GCACCTCCCTCGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCGCTGGCCCTCCAT 213
Db 441 AlaProProProProProProPro-----ProProProProPro 453
Qy 214 CCAGAAAACCTATGACCTCACCCCTACCTGGA---GCATCAACTCCCGCAGCTTAGCTGG 270
Db 454 ProProProLeuProProProProProProProLeuAlaSerLeuSerHisCysGlySer 473
Qy 271 GACCTACCTGAACCTACCTGGGGCCCCCTTT-----CAACGAGCCTGA 312
Db 474 GlnAlaSerProProProGlyThrProLeuAlaSerThrProSerSerLysProSerVal 493
Qy 313 CTTCAATCTCTCGACTGCGGGGAGAACTCTGCCAGGCG-----CACGTCACATT 366
Db 494 LeuProSerProSerAlaGlyAlaProAlaSerAlaGluThrProLeuAsnProGluLeu 513
Qy 367 GGA-----AGTGTGGCGAAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGAGGC 420
Db 514 GlyAspSerSerAlaSerGluProGlyLeuGlnAlaAlaSerGlnPro----- 529
Qy 421 GTACAGTACCT---CCTGTGTTACTTGTGCTGGGCTCAACCGTCAAGCTGCGCAGCTGA 477
Db 530 AlaGluSerProThrProGlnGlyLeuValLeuGlyProProAla----- 544
Qy 478 ACTCGAGTAGCTTGGCCGACCTTCTGTACAGCTCCAGGCGCTGCTGGGCGAGCATTCG 537
Db 545 ---ProProProProProProLeu-----ProSerGlyProAla-TyrAlaSerAl 560
Qy 538 AGGTGTCTATGGCGAGCTTGGCTACCCACTGCCAGGCTCTGCCAGGAGCTGAGCCAGC 597
Db 560 a---LeuProProProProGlyProProProProProProProProProProPro 579
Qy 598 CTGGGCGCTGGCCCT 613
Db 579 oProProProProPro 584

RESULT 6
US-10-001-873-50
; Sequence 50, Application US/10001873
; Patent No. US20020160388A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and
; FILE REFERENCE: DEX-0275
; CURRENT APPLICATION NUMBER: US/10/001,873
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,055
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,496
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-873-50
```



```

      |||||
Db 891 ArgGlyAlaArgGlyProThrGlyLysProGlyProLysGlyThrSerGlyGlyAspGly 910
      |||||
Qy 91 CCCCATGGACCTCCGAGC-----AGGGAGCTCTGGGGGATGTAGCTTCCT 138
      |||||
Db 911 ProProGlyProProGlyGluArgGlyProGlnGlyProGlnGlyProValGlyPhePro 930
      |||||
Qy 139 ATGCAGGTGCTGTGCACCTCCCTCAGTGCCAGCT----- 175
      |||||
Db 931 -----GlyProLysGlyPro-ProGlyProProGlyArgMetGlyCysProGlyHisPr 948
      |||||
Qy 176 -----CTTAATGCACAGGAGATCCAGGCCCTCGCCCTC 210
      |||||
Db 948 oGlyGlnArgGlyGluThrGlyPheGlnGlyThrGlyProProGlyProGlyGlyVa 968
      |||||
Qy 211 CATC-----CAGAAACCTATGACCTCACCTGCTACCTGAGCATCACTCCGAGCTT 264
      |||||
Db 968 lValGlyProGlnGlyProThrGlyGluThrGlyProLysGlyGluArg----- 984
      |||||
Qy 265 AGCTGGACCTACCTGAACCTACTGGGGCCCTTTCACAGAGCCTGACTTCAATCTCTCC 324
      |||||
Db 985 -----GlyTyProGlyProPro-----GlyProPr 993
      |||||
Qy 325 T-----CGACTGGGGGAGAACTCTGCCAGGCG 354
      |||||
Db 993 oGlyGlnGlnGlyLeuProGlyAlaAlaGlyLysGlyGlyAla-LysGlyAspProGlyP 1013
      |||||
Qy 355 CACGCTCA---ACTTGAAGTGTGGGAAGCCTCAATGACAGCGCTGC---GGCTGACCCA 408
      |||||
Db 1013 roGlnGlyLeuSerGlyLysAspGlyProAlaGlyLeuArgGlyPheProGly-GluArg 1032
      |||||
Qy 409 GAACATCAGGGGTACAGTCACTCTGTGTACTTGTGGTGGCTCAACGTCAGGCTGC 468
      |||||
Db 1033 GlyLeuProGlyAlaGlnGlyAlaProGlyLeuLysGlyGlyGlyProGlnGly--- 1051
      |||||
Qy 469 CACAGCTGAACCTCCGAGCTAGCTGSCCCACTTCTGTACCAGCTCCAGGCTGTGG 528
      |||||
Db 1052 -----ProProGlyPro-ValGI 1057
      |||||
Qy 529 CAGC-----ATTGAGGTGTATGGCGAGCGCTTGGCTACCC 564
      |||||
Db 1057 ySerProGlyGluArgGlySerAlaGlyThrAlaGlyProLysGlyLeuArgGlyArgPr 1077
      |||||
Qy 565 ACTGCCACGCTCTGCCAGGAGCTAGCCAGC-----TGGGCCCT 607
      |||||
Db 1077 oGlyProGlnGlyProProGly-----ProAlaGlyGluLysGlyAlaProGlyGly 1095
      |||||
Qy 608 -----GGCCTGTGCCACAGTGACTTCTCTCCAG 634
      |||||
Db 1095 sGlyProGlnGlyProAlaGlyArgAspGlyValGln 1107
      |||||

RESULT 10
US-09-770-689A-4
; Sequence 4, Application US/09770689A
; Patent No. US2002011517A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 726
; TYPE: PRT
; ORGANISM: HUMAN
US-09-770-689A-4
Alignment Scores:
Pred. No.: 0.0211 Length: 726
```

```

Score: 123.00 Matches: 54
Percent Similarity: 36.63% Conservative: 20
Best Local Similarity: 26.73% Mismatches: 72
Query Match: 8.21% Indels: 56
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-770-689A-4 (1-726)
Qy 31 CGCCTCCCTCCACTCCGCCAGCCTCTGGGAGAGAGCGCGCGCGCGCGGCC---GGCCCC 87
      |||||
Db 518 ArgSerProSerProProThrGlnHisThrGlyGlnProProGlyGlnProSerAlaPro 537
      |||||
Qy 88 CAGCCCCATGACCTCCGAGCAGG-----GGACTCTGTGGGGGATGTAGCTTGCCTATG 141
      |||||
Db 538 SerGlnLeuSerAlaProArgArgTySerSerSerLeuSerProIleGlnAlaProAsn 557
      |||||
Qy 142 CACGGTCTGTGGCACCTCCCTGAGTCCAGTCCAGCTTCTTAATCCACAGGAGATCCAGG--- 198
      |||||
Db 558 HisProProGlnProProThrGlnAlaThrProLeuMetHisThrLysProAsnSer 577
      |||||
Qy 199 -----CCTCGGCCCTC-----CATCCAGAAAC 222
      |||||
Db 578 GlnGlyProProAsnProMetAlaLeuProSerGluHisGlyLeuGluGlnProSerHis 597
      |||||
Qy 223 CTATGACCTCACCCGCTACTCTGGAGCATCAACTCCGACGTTAGCTGGACCTACCTGAA 282
      |||||
Db 598 ThrProGlnThrProThrProProSerThrProProLeuGlyLysGlnAsnProSer 617
      |||||
Qy 283 CTACTCGGGGCCCTTTCACAGGAGCTGACTTCAATCTCTCGACTGGGGGAGAAAC 342
      |||||
Db 618 LeuPro---AlaPro-----GlnThrLeuAlaGlyGlyAsnProGlu 630
      |||||
Qy 343 TCTGCCAGGCGCACGCTCAACTTGAAGTGTGGGAAGCCTCAATGACAGCTGCGGCT 402
      |||||
Db 631 ThrAlaGlnProHisAla-----GlyThrLeuProArgProArg--- 643
      |||||
Qy 403 GACCCAGAACTATGAGGCGTACATCCTCTGTGTACTTGTGCTGGCCTCAACCGTCA 462
      |||||
Db 644 -----ProVal-----ProLysProArg 649
      |||||
Qy 463 GGCTGCCACAGCTGAATCCGACGCTAGCTGGCCACCTTCTGTACCAGCTCCAGGCT 522
      |||||
Db 650 AsnArgProSerValProPro-----ProProGlnPro 660
      |||||
Qy 523 GCTGGGAGCATTCAGGTGTATGCGAGCGCTTGGCTACCC---ACTGCCCGACCTCT 579
      |||||
Db 661 ProGlyValHisSerAla-----GlyAspSerSerLeuThrAsnThrAlaProThrAla 678
      |||||
Qy 580 GCCAGG 585
      |||||
Db 679 SerLys 680

RESULT 11
US-09-770-689A-2
; Sequence 2, Application US/09770689A
; Patent No. US2002011517A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 803
; TYPE: PRT
; ORGANISM: HUMAN
US-09-770-689A-2
Alignment Scores:
```


Pred. No.: 0.0214 Length: 803
Score: 123.00 Matches: 54
Percent Similarity: 36.53% Conservatives: 20
Best Local Similarity: 26.73% Mismatches: 72
Query Match: 8.21% Indels: 56
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-770-689A-2 (1-803)

QY 31 CGCCCTCCCACTCCGCGCAGCTCTGGAGAGAGCGCGCGCGCGCGCC---GGCCCC 87
Db 595 ArgSerProSerProThrGlnHisThrGlnProProGlyGlnProSerAlaPro 614
QY 88 CAGCCCCATGACCTCCGAGCAGG-----GGACTCGTGGGGATGTTAGCTGCCTATG 141
Db 615 SerGlnLeuSerAlaProArgArgTyrSerSerSerLeuSerProIleGlnAlaProAsn 634
QY 142 CACGGTGTGTGGCACCTCCCTGCGAGTGCAGCTCTTAATCGCACAGGAGATCCAGG--- 198
Db 635 HisProProGlnProProThrGlnAlaThrProLeuMetHisThrLysProAsnSer 654
QY 199 -----CCCTGGCCCTC-----CATCAGAAAC 222
Db 655 GlnGlyProProAsnProMetAlaLeuProSerGluHisGlyLeuGlnProSerHis 674
QY 223 CTATGACCTCACCGCTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACCTACCTGAA 282
Db 675 ThrProProGlnThrProThrProProSerThrProProLeuGlyLysGlnAsnProSer 694
QY 283 CTACCTGGGGCCCCCTTTCAACAGCGCTGACCTCAATCCTCCTCGACTGGGGGCGAGAAC 342
Db 695 LeuPro---AlaPro-----GlnThrLeuAlaGlyGlyAsnProGlu 707
QY 343 TCTGCCAGGCGCCAGCTCACTTGAAGTGTGGGAGCGCTCAATGACAGGCTGGGCT 402
Db 708 ThrAlaGlnProHisAla-----GlyThrLeuProArgProArg----- 720
QY 403 GACCAGAACTAGAGCGGTACAGTCACCTCTGTGTGTACTTGGCGGCTCAACCGTCA 462
Db 721 -----ProVal-----ProLysProArg 726
QY 463 GGCTGCCACAGCTGAATCCGAGTAGCTGGCGCCACTTCTGTACGAGCTCCAGGGCCT 522
Db 727 AsnArgProSerValProPro-----ProProGlnPro 737
QY 523 GCTGGGAGCATGCGAGTGTATGGCGAGCGTGGCTACCC---ACTGCCCCAGCCTCT 579
Db 738 ProGlyValHisSerAla-----GlyAspSerSerLeuThrAsnThrAlaProThrAla 755
QY 580 GCCAGG 585
Db 756 SerLys 757

RESULT 12
US-09-816-860A-2
; Sequence 2, Application US/09816860A
; Patent No. US20020081651A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 26649, A No. US20020081651A1 Human GTPase Activating Molecule
; FILE REFERENCE: MNT-133
; CURRENT APPLICATION NUMBER: US/09/816, 860A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/191,859
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 881
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-816-860A-2

Alignment Scores:
Pred. No.: 0.0218 Length: 881
Score: 123.00 Matches: 54
Percent Similarity: 36.63% Conservatives: 20
Best Local Similarity: 26.73% Mismatches: 72
Query Match: 8.21% Indels: 56
DB: 10 Gaps: 12

US-09-931-704-4 (1-819) x US-09-816-860A-2 (1-881)

QY 31 CGCCCTCCCACTCCGCGCAGCTCTGGAGAGAGCGCGCGCGCGCC---GGCCCC 87
Db 673 ArgSerProSerProThrGlnHisThrGlnProProGlyGlnProSerAlaPro 692
QY 88 CAGCCCCATGACCTCCGAGCAGG-----GGACTCGTGGGGATGTTAGCTGCCTATG 141
Db 693 SerGlnLeuSerAlaProArgArgTyrSerSerSerLeuSerProIleGlnAlaProAsn 712
QY 142 CACGGTGTGTGGCACCTCCCTGCGAGTGCAGCTCTTAATCGCACAGGAGATCCAGG--- 198
Db 713 HisProProGlnProProThrGlnAlaThrProLeuMetHisThrLysProAsnSer 732
QY 199 -----CCCTGGCCCTC-----CATCAGAAAC 222
Db 733 GlnGlyProProAsnProMetAlaLeuProSerGluHisGlyLeuGlnProSerHis 752
QY 223 CTATGACCTCACCGCTACCTGGAGCATCAACTCCGAGCTTAGCTGGGACCTACCTGAA 282
Db 753 ThrProProGlnThrProThrProProSerThrProProLeuGlyLysGlnAsnProSer 772
QY 283 CTACCTGGGGCCCCCTTTCAACAGCGCTGACCTCAATCCTCCTCGACTGGGGGCGAGAAC 342
Db 773 LeuPro---AlaPro-----GlnThrLeuAlaGlyGlyAsnProGlu 785
QY 343 TCTGCCAGGCGCCAGCTCACTTGAAGTGTGGGAGCGCTCAATGACAGGCTGGGCT 402
Db 786 ThrAlaGlnProHisAla-----GlyThrLeuProArgProArg----- 798
QY 403 GACCAGAACTAGAGCGGTACAGTCACCTCTGTGTGTACTTGGCGGCTCAACCGTCA 462
Db 799 -----ProVal-----ProLysProArg 804
QY 463 GGCTGCCACAGCTGAATCCGAGTAGCTGGCGCCACTTCTGTACGAGCTCCAGGGCCT 522
Db 805 AsnArgProSerValProPro-----ProProGlnPro 815
QY 523 GCTGGGAGCATGCGAGTGTATGGCGAGCGCTGGCTACCC---ACTGCCCCAGCCTCT 579
Db 816 ProGlyValHisSerAla-----GlyAspSerSerLeuThrAsnThrAlaProThrAla 833
QY 580 GCCAGG 585
Db 834 SerLys 835

RESULT 13
US-10-086-464-17
; Sequence 17, Application US/10086464
; Publication No. US20020199218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762-A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13

```
; *NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-086-464-17

Alignment Scores:
Pred. No.:          0.0278          Length:          731
Score:             121.50          Matches:         59
Percent Similarity: 36.73%          Conservatives:    13
Best Local Similarity: 30.10%          Mismatches:     93
Query Match:       8.11%             Indels:         32
DB:                9                Gaps:           6

US-09-931-704-4 (1-819) x US-10-086-464-17 (1-731)
QY 34 CCTCCACTCCGCCAGCTCTGGGAGAGGAGCGCGCGCGCGCGCGCGCGCGCG 93
Db 32 ProProAlaProProLeuSerProLeuProProProProLeuSerSerProProLeuPro 51
QY 94 CATGACCTCCG---AGCAGGGGACTCGTGGGGGATGTTAGCTTGCTATGCACGGTGCT 150
Db 52 SerProProProLeuSerAlaProThraAlaSerProProProProLeuProValGluSerPro 71
QY 151 GTGGCACTCCCTGCAGTCCAGCTCTTAATCGCAGGAGATCCAGGCGCTCGCCCTC 210
Db 72 ProSerProProIleGluSerProProProProProProProLeuGluSerProProProPro 90
QY 211 CATCAGAAAACCTATGACCTCACCGCTACCTGGAGGATCAACTCGGAGCTTAGCTGG 270
Db 91 ---ProLeuGluSerProSerProProSerProHisValSerAlaProSerGlySerPro 109
QY 271 GACCTACCTGAACCTACCTGGGCGCCCTT-----TCAACGAGCGCTGACTTCAATCCTC 323
Db 110 ProLeuProPheLeuProAlaLysProSerProProProSerSerProProSerGluThr 129
QY 324 CTCGACTGGGGCAGAACTCTGCCAGGCGCCACGTCGAACCTTGAAGTGTGGCGAAGCC 383
Db 130 ValProProGlyAsnThrIleSerProProProProProProSerProSerGluSerThrPr 149
QY 384 TCAATGACAGCTGGGCTGAGCCAGCACTATGAGCGGTACAGTCACTCTGTGTACT 443
Db 149 oProValAsnThrAlaSerProPro-----ProSerProProArgArgAr 165
QY 444 TCGGTGGCTCAACGCTCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCACTTCT 503
Db 165 sSerGlyProLysProSerPhe-----ProProProIleAs 177
QY 504 GTACAGGCTCCAGGCGCTGCTGGGAGCATTTGACAGGTGTATGGCGAGCGCTTGGCTACC 563
Db 177 nSerSerProProAsnProSer-----ProProProProProProProProProPro 185
QY 564 CACTGCCCGAGCTTGCAGGAGCTGACCGACCTGGCGCCCT 607
Db 185 oAsnThrProSer-LeuProGluThrSerProProProLysPro 199

RESULT 14
US-09-843-676-217
; Sequence 217, Application US/09843676
; Patent No. US20020164786A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
;             Lingner, Joachim
;             Nakamura, Toru
;             Chapman, Karen B.
;             Morin, Gregg B.
;             Harley, Calvin
;             Andrews, William H.
; TITLE OF INVENTION: No. US20020164786A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/843,676
; FILING DATE: 26-Apr-2001
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US/08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US/08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US/08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 217:
US-09-843-676-217

Alignment Scores:
Pred. No.:          0.0321          Length:          1003
Score:             121.00          Matches:         81
Percent Similarity: 31.17%          Conservatives:    20
Best Local Similarity: 25.00%          Mismatches:     104
Query Match:       8.08%             Indels:         119
DB:                9                Gaps:           18

US-09-931-704-4 (1-819) x US-09-843-676-217 (1-1003)
QY 23 CCGCGGCTCGCCCTCCCACTCCGCGAGGAGCGCGCGCGCGCGCGCGCGCG 82
Db 110 ProGlyValAlaLeu-----ProLeuSerArgSerGlyArgProLeuGlyArg 125
QY 83 GCCCCAGCCCCATGACCTCCGAGCAGGG-----GACTCGTGGGGATGTAGTTGCTGC 136
Db 126 GlyProGlyPro-----ThrArgAlaGlyArgLeuAspArgValThrValSerVal 143
QY 137 CTATGACGCTGTGTGGCACCTCCCTGACAGTGCA-----GCTCTTAATCGCACAGGATCCA 172
Db 144 TrpCys-----HisLeuProAspProProLysLysProProLeuTrpArgVal 159
QY 173 -----GCCCCCTCCATCCAGAAACC 223
Db 160 ArgSerLeuAlaArgAlaThrProThrHisProTrpAlaAlaSerThrArgAlaPro 179
QY 197 GGCCCT-----GGCCCTCCATCCAGAAACC 223
Db 180 HisProHisArgGlyHisHisValLeuGlyHisAlaLeuSerProGlyValArgArgAsp 199
QY 224 TATGACCTCACCGCTACCTGGAGCATCACTCCGACGCTTACTGCTGGAGCTTACCTGAAC 283
Db -----
```

```
Db 200 GlnAlaLeuProLeuLeuLeuArgGln***HisCys***ProProSerTyrSerIle 219
Qy 284 TACCTGGGGCCCTTTCAACGAGCGTCACTTCAATCCTCTCGAC----- 329
Db 220 TyrLeuArgProSerLeu-ThrGlyValArgGluValArgGlyAsp***LeuSerGlySe 239
Qy 330 -----TGGGGGCGCAAACTC-----TGCCAGGG---CCACGGTCAACTTGGAA 370
Db 239 rArgProTrpMetProGlyPheProAlaGlyCysProAlaCysProSer***ThrGlyLy 259
Qy 371 GTGTGGC-----GAAGCCTCAATACACAGCTGGCG 400
Db 259 sCysGlyProCysPheTrpSerCysLeuGlyThrThrArgSerAlaProThrGlyCysSe 279
Qy 401 CTGACCCAGAACTATGAGCGGTACAGTCACCTCCTGTGTACTTGGTGGCCTCAACCGT 460
Db 279 rSerArgArgThrAlaArg-----CysGluLe 288
Qy 461 CAGGTGGCCACAGCTGAACCTCGAGCTAGCTGGCCCACTTCTGTACCA-----GCCTC 514
Db 288 uArgSerProGln-----GlnProValSerValProGlyArgSerPr 302
Qy 515 CAGGCGCTGTGGGAGCAGATTGAGGTGTATGGGAGCGCTGGCTACCCACTGCCCCAG 574
Db 302 oArgAlaLeuTrpArgProPro-----ArgArgAsnThrAspProArgAr 318
Qy 575 CTTCTGCCAGGAGCTGAGCGAGCTGGGCCCTGGGCCCTGGCCACAGTCACTTCTCCAG 634
Db 318 GluValGlnLeuLeuArgGln-----HisSerSe 328
Qy 635 AAGATGGATGACTTCTGGCTGTGTAAGGAGCTGC-----AGACTGGCTA 679
Db 328 rProTrpGlnValTyrGlyPheValArgAlaCysLeuArgGluValProGlyLe 348
Qy 680 T-----GGGTTCAGCCCAAGGACTTCAAC 703
Db 348 uTrpGlySerArgHisAsnGluArgPheLeuArgAsnThrLysLysPheIleSerLe 368
Qy 704 CGGCTTAAGAGAAGATGAGCTCCAGCTTCAGTTCAGTCCCTGCGACTTGGAGGCACAT 763
Db 368 uGlyLysHisAlaLysLeuSerLeuGlnGluLeu-----ThrTrpLysMetSe 384
Qy 764 GTTTCCTGAC 773
Db 384 rValargasp 387

RESULT 15
US-09-438-486-217
; Sequence 217, Application US/09438486
; Publication No. US20030009019A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. US20030009019A1el Telomerase
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/438,486
; FILING DATE: 12-NOV-1999
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002931US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-438-486-217

Alignment Scores:
Pred. No.: 0.0321 Length: 1003
Score: 121.00 Matches: 81
Percent Similarity: 31.17% Conservative: 20
Best Local Similarity: 25.00% Mismatches: 104
Query Match: 8.08% Indels: 119
DB: 9 Gaps: 18

US-09-931-704-4 (1-819) x US-09-438-486-217 (1-1003)
Qy 23 CCGGGCTCGCCCTCCACTCCGAGCGAGGAGCGCGCGCGCGCGCG 82
Db 110 ProGlyValAlaLeu-----ProLeuSerArgSerGlyArgProLeuGlyArg 125
Qy 83 GCGCCAGCGCCATGACCTCCGAGCAGGG-----GACTCGTGGGGATGTAGCTTGC 136
Db 126 GlyProGlyPro-----ThrArgAlaGlyArgLeuAspArgValThrValSerVal 143
Qy 137 CTATGCACGGTGTGTGGCACTCCCTGCACTGCCA----- 172
Db 144 TrpCys-----HisLeuProAspProLysLysProLeuTrpArgVal 159
Qy 173 -----GCTCTTAATCGCACGAGATCCA 196
Db 160 ArgSerLeuAlaArgAlaThrProThrHisProTrpAlaAlaSerThrArgAlaPro 179
Qy 197 GGCCCT-----GGCCCTCCATCCAGAAACC 223
Db 180 HisProHisArgGlyHisHisValLeuGlyHisAlaLeuSerProGlyValArgArgAsp 199
Qy 224 TATGACTTCACTCCCTACCTGAGAGATCACTCCGCACTTAGCTGGGACCTACCTGAAC 283
Db 200 GlnAlaLeuProLeuLeuArgGln***HisCys***ProProSerTyrSerIle 219
Qy 284 TACCTGGGGCCCTTTCAACGAGCGTCACTTCAATCCTCTCGAC----- 329
Db 220 TyrLeuArgProSerLeu-ThrGlyValArgGluValArgGlyAsp***LeuSerGlySe 239
```

330	Qy	-----TGGGGCAGAAATCTC-----TGCCACAGG-----CCACGGCTCAACTTGGAA	370
239	Db	rArgProTirpMetPirProGlyPheProAlaGlyCysPsoAlaCysProSer**ThrGlyly	259
371	Qy	GTGTGGC-----GAAGCTCAATCACAGGCTGGCG	400
259	Db	sCysGlyProCysPheTirpSerCysLeuglyThrThrArgSerAlaProThrGlyCys	279
401	Qy	CTGACCAGAACTATGAGGGCTAGTACGTACCTCTGTATTCTGCTGGCCCTCAACCGT	460
279	Db	rSerArgArgThrAlaarg-----CysGluLe	288
461	Qy	CAGGCTCCACAGCTGAACCTCCAGAGTAGCTGGCCACCTTCTGTACCA-----GCCTC	514
288	Db	uArgSerProGln-----GlnProValSerValProGlyArgSerPr	302
515	Qy	CAGGGCTGTGTGGCAGCATTTGCAGGTGTCTATGGCAGCGCTTGGCTACCACTGCCCG	574
302	Db	oArgAlaLeuTirpArgProPro-----ArgArgAsnThrAspProArgAr	318
575	Qy	CCTCTGCCAGGGACTGAGCCAGAGCTGGGCCCTTGGCCCTGCCACAGTGACTTCTCTCCAG	634
318	Db	gLeuValGlnLeuLeuArgGln-----HisSerSe	328
635	Qy	AAGATGATGACTTCTGGCTGTCTGAGGAGCTGC-----AGACCTGGCTTA	679
328	Db	rProTirpGlnValTyrGlyPheValArgAlaCysLeuArgLeuValProGlyLe	348
680	Qy	T-----GGCGTTCAGCCCAAGAGACTTCAAC	703
348	Db	uTirpGlySerArgHisasnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLe	368
704	Qy	CGGCTTAAGAAGAAGATGCAGCCTCCAGACGCTTCAGTACCCTGCAGCTTGGAGGCACAT	763
368	Db	uGlyLysHisAlaLysLeuSerLeuGlnGluLeu-----ThrTirpLysMetSe	384
764	Qy	GGTTTCTGC	773
384	Db	rValArgAsp	387

Search completed: January 27, 2003, 17:23:06
Job time : 17.8648 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 27, 2003, 15:50:02 ; Search time 8.30852 Seconds
(without alignments)
5800.643 Million cell updates/sec

Title: US-09-931-704-4

Perfect score: 1498

Sequence: 1 tattattaaagcttcgcgg.....aggccacagtcagctgtgctt 819

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09931704/runat_27012003_154127_3671/app_query.fasta_1.7189
-DB=Issued Patents AA -QWTF=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPFCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931704@cgn 1.89 @runat_27012003_154127_3671 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WAEV TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	81.7	225	1	US-08-792-019B-5
2	1224	81.7	225	3	US-08-988-819-5
3	1224	81.7	225	4	US-09-016-534-5
4	1193	79.6	225	1	US-08-792-019B-2
5	1193	79.6	225	3	US-09-106-182-2
6	1193	79.6	225	4	US-08-988-819-2
7	1193	79.6	225	3	US-09-016-534-2
8	146	9.7	330	1	US-08-642-255-32
9	146	9.7	408	1	US-07-609-716-65
10	146	9.7	408	4	US-08-475-411A-65
11	146	9.7	408	4	US-08-478-029A-65
12	140.5	9.4	1064	1	US-08-642-255-62

13	139	9.3	684	1	US-08-555-669-12	Sequence 12, Appl	
14	139	9.3	684	3	US-09-073-663-12	Sequence 12, Appl	
15	136	9.1	682	1	US-08-642-255-126	Sequence 126, App	
16	136	9.1	682	1	US-08-397-633A-36	Sequence 36, Appl	
17	136	9.1	960	4	US-09-219-849-5	Sequence 5, Appl	
18	135.5	9.0	252	1	US-08-642-255-61	Sequence 61, Appl	
19	135	9.0	357	1	US-07-609-716-66	Sequence 66, Appl	
20	135	9.0	357	1	US-08-642-255-33	Sequence 33, Appl	
21	135	9.0	357	4	US-08-475-411A-66	Sequence 66, Appl	
22	135	9.0	357	4	US-08-478-029A-66	Sequence 66, Appl	
23	135	9.2	493	4	US-08-556-978B-59	Sequence 59, Appl	
24	135	9.0	762	1	US-08-642-255-120	Sequence 120, App	
25	135	9.0	762	1	US-08-397-633A-31	Sequence 31, Appl	
c	26	135	870	2	US-09-010-928B-2	Sequence 2, Appl	
c	27	134	9.2	529	4	US-08-247-806-2	Sequence 2, Appl
c	28	134	9.2	595	1	US-08-425-069-4	Sequence 4, Appl
c	29	134	9.2	595	2	US-08-317-844B-4	Sequence 4, Appl
30	133.5	8.9	829	1	US-08-642-255-132	Sequence 132, App	
31	133.5	8.9	829	1	US-08-397-633A-53	Sequence 53, Appl	
32	132	8.8	761	2	US-08-707-237A-84	Sequence 84, Appl	
33	132	8.8	762	1	US-08-642-255-114	Sequence 114, App	
34	132	8.8	762	1	US-08-397-633A-26	Sequence 26, Appl	
35	131.5	8.8	504	4	US-09-219-849-3	Sequence 3, Appl	
36	131.5	8.8	561	1	US-08-642-255-52	Sequence 52, Appl	
37	131.5	8.8	720	4	US-09-219-849-4	Sequence 4, Appl	
38	131.5	8.8	777	1	US-08-642-255-53	Sequence 53, Appl	
c	39	130	8.9	1077	1	US-07-972-032-82	Sequence 82, Appl
c	40	130	8.9	1077	1	US-08-642-255-95	Sequence 95, Appl
41	129.5	8.6	417	1	US-08-175-155-69	Sequence 69, Appl	
42	129.5	8.6	417	1	US-08-477-509B-104	Sequence 104, App	
43	129.5	8.6	417	1	US-08-642-255-102	Sequence 102, App	
44	129.5	8.6	417	2	US-08-707-237A-76	Sequence 76, Appl	
45	129.5	8.6	417	3	US-08-482-085B-104	Sequence 104, App	

ALIGNMENTS

RESULT 1

US-08-792-019B-5
; Sequence 5, Application US/08792019B
; Patent No. 5741772
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: 1840 DEHAVILLAND DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,019B
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: COOK, ROBERT R.
; REGISTRATION NUMBER: 31,602
; REFERENCE/DOCKET NUMBER: A-442
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-792-019B-5

Alignment Scores:
Pred. No.: 3.7e-108 Length: 225
Score: 1224.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.71% Indels: 0
DB: 1 Gaps: 0

US-09-931-704-4 (1-819) x US-08-792-019B-5 (1-225)

QY 95 ATGGACCTCCGAGCAGGAGCTGCTGGGATGTTAGCTTGCTATGCAGCGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyValLeuTrp 20

QY 155 CACCTCCCTGCGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCCCTGGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACCTCACCCGCTACCTGGAGCATCACTCCGAGCTTACTGGGACC 274
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAACCTACCTGGGCCCCCTTCAACGAGCTGACTTCAATCCTCCTCGACTGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAAACTCTGCCCGAGGCCACCGTCAACTTGGAGTGTGGCGAAGCTCAATGACAG 394
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100

QY 395 CTGCGGTGACCCAGAACTATGAGCGGTACAGTCACCTCCTGTGTACTTGGTGGCCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGAGTGCCTGGCCCTCTGTATCCAGCCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGCCCTGCTGGGCGAGCATTCAGGTGCTCATGGCGAGCTTGGCTACGCCCCAG 574
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CCTCTGCCAGGAGTGCAGCCAGCTGGGCCCTGGCCCTGCCACAGTACTTCTCTCCAG 634
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGGATGACTTCTGGCTCTGAAGAGCTGCAGACCTGGCTATGGCGTTGAGCCAG 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAAAGATGCAGCCTCCAGCAGCTTCAGTCACCTGCACCTG 754
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220

QY 755 GAGGCACATGGTTTC 769
Db 221 GluAlaHisGlyPhe 225

RESULT 2
US-08-988-819-5
; Sequence 5, Application US/08988819
; Patent No. 6054294
; GENERAL INFORMATION:
; APPLICANT: CHANG, MING-SHI
; TITLE OF INVENTION: NEUROTROPHIC FACTOR NNT-1
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMGEN INC.
; STREET: ONE AMGEN CENTER DRIVE
; CITY: THOUSAND OAKS
; STATE: CA
; COUNTRY: USA
; ZIP: 91320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/988,819
FILING DATE: 12-DEC-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/792,019
FILING DATE: 03-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: COOK, ROBERT R.
REGISTRATION NUMBER: 31,602
REFERENCE/DOCKET NUMBER: A-442A
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-988-819-5

Alignment Scores:
Pred. No.: 3.7e-108 Length: 225
Score: 1224.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.71% Indels: 0
DB: 3 Gaps: 0

US-09-931-704-4 (1-819) x US-08-988-819-5 (1-225)

QY 95 ATGGACCTCCGAGCAGGAGCTGCTGGGATGTTAGCTTGCTATGCAGCGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuGlyValLeuTrp 20

QY 155 CACCTCCCTGCGAGTGCAGCTCTTAATCGCACAGGAGATCCAGGCCCTGGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerIle 40

QY 215 CAGAAAACCTATGACCTCACCCGCTACCTGGAGCATCACTCCGAGCTTACTGGGACC 274
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60

QY 275 TACCTGAACCTACCTGGGCCCCCTTCAACGAGCTGACTTCAATCCTCCTCGACTGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80

QY 335 GCAGAAACTCTGCCCGAGGCCACCGTCAACTTGGAGTGTGGCGAAGCTCAATGACAG 394
Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100

QY 395 CTGCGGTGACCCAGAACTATGAGCGGTACAGTCACCTCCTGTGTACTTGGTGGCCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120

QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGAGTGCCTGGCCCTCTGTATCCAGCCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140

QY 515 CAGGCCCTGCTGGGCGAGCATTCAGGTGCTCATGGCGAGCTTGGCTACGCCCCAG 574
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160

QY 575 CCTCTGCCAGGAGTGCAGCCAGCTGGGCCCTGGCCCTGCCACAGTACTTCTCTCCAG 634
Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180

QY 635 AAGATGGATGACTTCTGGCTCTGAAGAGCTGCAGACCTGGCTATGGCGTTGAGCCAG 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200

QY 695 GACTTCAACCGGCTTAAGAAGAGATGACGCTCCAGCAGCTTCAGTCAACCTGCACTTG 754
 |||||
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 |||||
 QY 755 GAGGCACATGGTTTC 769
 |||||
 Db 221 GluAlaHisGlyPhe 225
 |||||

RESULT 3

US-09-016-534-5
 ; Sequence 5, Application US/09016534
 ; Patent No. 6143874
 ; GENERAL INFORMATION:
 ; APPLICANT: CHANG, MING-SHI
 ; APPLICANT: ELLIOTT, GARY S.
 ; APPLICANT: SARMIENTO, ULLA
 ; APPLICANT: SENALDI, GIORGIO
 ; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: AMGEN INC.
 ; STREET: ONE AMGEN CENTER
 ; CITY: THOUSAND OAKS
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 91320
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/016,534
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/792,019
 ; FILING DATE: 03-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: COOK, ROBERT R.
 ; REGISTRATION NUMBER: 31,602
 ; REFERENCE/DOCKET NUMBER: A-442B
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 225 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-016-534-5

Alignment Scores:
 Pred. No.: 3,7e-108 Length: 225
 Score: 1224.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.71% Indels: 0
 DB: 4 Gaps: 0

US-09-931-704-4 (1-819) x US-09-016-534-5 (1-225)

QY 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTAGCTTGCTATGACCGTGTGTGG 154
 |||||
 Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
 |||||
 QY 155 CACCTCCCTGAGTCCAGCTCTTAATCGCACAGAGATCCAGGCCCTGGCCCTCCATC 214
 |||||
 Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProSerIle 40
 |||||
 QY 215 CAGAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
 |||||
 Db 41 GlnLysThrThrAspLeuThrArgTrpLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
 |||||

QY 275 TACCTGAACCTACCTGGGGCCCCCTTTCAACGAGCCTGACTTCAATCTCTCGACTGGGG 334
 |||||
 Db 61 TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly 80
 |||||
 QY 335 GCAGAAACTCTGCCAGGCCACGGTCAACTTGGAAAGTGTGGCGAAGCCTCAATGACAGG 394
 |||||
 Db 81 AlaGluThrLeuProArgAlaThrValAsnLeuGluValTrpArgSerLeuAsnAspArg 100
 |||||
 QY 395 CTGGCGCTGACCCAGCACTATGAGCGGTACAGTCACCTCTGTGTACTTGGCTGGCCTC 454
 |||||
 Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
 |||||
 QY 455 AACCGTCAGGCTGCCACAGCTGAACCTCCGACGTAGCTGGCCCTCTTGTGTACAGCCTC 514
 |||||
 Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
 |||||
 QY 515 CAGGGCTCTGGCGAGCATTTGAGGTTCATGGCGACGCTTGGCTTACCCACTGCCCCAG 574
 |||||
 Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaThrLeuGlyTyrProLeuProGln 160
 |||||
 QY 575 CTCTGCCAGGACTGAGCCAGCCTGGCGCCCTGGCCCTGCCACAGTGACTTCTCTCCAG 634
 |||||
 Db 161 ProLeuProGlyThrGluProAlaTrpAlaProGlyProAlaHisSerAspPheLeuGln 180
 |||||
 QY 635 AAGATGGATGACTTCTGGCTGCTGAAGGAGCTGCAGACCTGGCTATGGGGTTTCAGCCAAG 694
 |||||
 Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
 |||||
 QY 695 GACTTCAACCGGCTTAAGAAGAGATGACGCTCCAGCAGCTTCAGTCAACCTGCACTTG 754
 |||||
 Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaSerValThrLeuHisLeu 220
 |||||
 QY 755 GAGGCACATGGTTTC 769
 |||||
 Db 221 GluAlaHisGlyPhe 225
 |||||

RESULT 4

US-08-792-019B-2
 ; Sequence 2, Application US/08792019B
 ; Patent No. 5741772
 ; GENERAL INFORMATION:
 ; APPLICANT: CHANG, MING-SHI
 ; TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: AMGEN INC.
 ; STREET: 1840 DEHAVILLAND DRIVE
 ; CITY: THOUSAND OAKS
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 91320
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/792,019B
 ; FILING DATE: 03-FEB-1997
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: COOK, ROBERT R.
 ; REGISTRATION NUMBER: 31,602
 ; REFERENCE/DOCKET NUMBER: A-442
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 225 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-792-019B-2

Alignment Scores:

Pred. No.: 3 22e-105 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 79.64% Indels: 0
DB: 1 Gaps: 0

US-09-931-704-4 (1-819) x US-08-792-019B-2 (1-225)

Qy 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCTATGCACGGTGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
Qy 155 CACCTCCCTGCGAGTCCAGCTCTTATTCGACAGGAGATCCAGCCCTGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40
Qy 215 CAGAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
Qy 275 TACCTGAACCTAGCTGGGGCCCCCTTCAACAGAGCTGACTTCAATCTCTCGACTGGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
Qy 335 CGAGAACTCTCCAGGCGCCAGCTCACTTGGAGTGTGGAGCTCAATGCACGG 394
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
Qy 395 CTGGCGCTGACCCAGAACTATGAGGGGTACAGTCACCTCTCTGTGTTACTTGGCGCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
Qy 455 AACCTGAGCTGCCAGAGCTGAACCTCCGAGCTAGCTGGGCCCTTCTGTACAGCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
Qy 515 CAGGCGCTGCTGGGAGCATTCAGGTGATGGCGACCTGGCTTACCCACTGCCCCAG 574
Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
Qy 575 CCTCTGCCAGGAGCTGAGCCAGCTGGGGCCCCCTGGCCCTGCCACAGTACTTCTCCAG 634
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
Qy 635 AGATGATGACTTCTGGCTGCTGAGGAGCTGACAGCTGCTGCTATGGCTTACGCCAAG 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
Qy 695 GACTTCAACCGGCTTAAGAAGAAGATCAGCTCCAGAGCTTCCAGTCCACCTGCACCTG 754
Db 201 AspPheAsnArgLeuLysLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
Qy 755 GAGGACATGGTTTC 769
Db 221 GlyAlaHisGlyPhe 225

RESULT 5
US-09-106-182-2
; Sequence 2, Application US/09106182
; Patent No. 6046035
; GENERAL INFORMATION:
; APPLICANT: Shi, Yangu
; APPLICANT: Ruben, Steve
; TITLE OF INVENTION: Cardiotrophin-Like Cytokine
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: #0850

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,182
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/051,053
FILING DATE: 30-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF385
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-182-2

Alignment Scores:
Pred. No.: 3 22e-105 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 79.64% Indels: 0
DB: 1 Gaps: 0

US-09-931-704-4 (1-819) x US-09-106-182-2 (1-225)

Qy 95 ATGGACCTCCGAGCGGAGCTCGTGGGGATGTTAGCTTGCTATGCACGGTGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
Qy 155 CACCTCCCTGCGAGTCCAGCTCTTATTCGACAGGAGATCCAGCCCTGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnArgThrGlyAspProGlyProGlyProSerile 40
Qy 215 CAGAAACCTATGACCTCACCGCTACCTGGAGCATCACTCCGACGCTTAGCTGGGACC 274
Db 41 GlnLysThrTyrAspLeuThrArgTyrLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
Qy 275 TACCTGAACCTAGCTGGGGCCCCCTTCAACAGAGCTGACTTCAATCTCTCGACTGGGG 334
Db 61 TyrLeuAsnTyrLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
Qy 335 CGAGAACTCTGCCAGGCGCCAGCTCACTTGGAGTGTGGAGCTCAATGCACGGTGTGTGG 394
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
Qy 395 CTGGCGCTGACCCAGAACTATGAGGGGTACAGTCACCTCTCTGTGTTACTTGGCGCTC 454
Db 101 LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu 120
Qy 455 AACCTGAGCTGCCAGAGCTGAACCTCCGAGCTAGCTGGGCCCTTCTGTACAGCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgSerLeuAlaHisPheCysThrSerLeu 140
Qy 515 CAGGCGCTGCTGGGAGCATTCAGGTGATGGCGACCTGGCTTACCCACTGCCCCAG 574
Db 141 GlnGlyLeuLeuGlySerileAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln 160
Qy 575 CCTCTGCCAGGAGCTGAGCCAGCTGGGGCCCCCTGGCCCTGCCACAGTACTTCTCCAG 634
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180

275	TACTGAACTACTGGGGCCCTTTCACAGAGCTGACTTCAATCTCTCGACTGGGG	333
61	TyrLeuAsnTyrLeuGlyProPheAsnGluProAspPheAsnProProArgLeuGly	80
335	GCAGAAACTCTGCCCAGGCGCAGGTCAACTTGGAGTGTGGCGAAGCTCAATCACAGG	394
81	AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys	100
395	CTGCGGCTGACCCAGAACTATGAGGCGTACAGTCACCTCCTGTGTACTTGGCGTGGCCTC	454
101	LeuArgLeuThrGlnAsnTyrGluAlaTyrSerHisLeuLeuCysTyrLeuArgGlyLeu	120
455	AACCGTCAGGTCGCCACAGCTGAATCCGAGCTAGCTGGGCCCACTTCTGTACCAAGCTTC	514
121	AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu	140
515	CAGGGCTCTGGGAGCAGTTCAGGTGCATGGCGAGCGCTTGGCTACCACTGCCCCAG	574
141	GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaAlaLeuGlyTyrProLeuProGln	160
575	CTCTGCCAGGACTGAGCAGCGCTGGGGCCCTGGCGCTGCCACAGTGAAGTTCCTCCAG	634
161	ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln	180
635	AGATGATGACTTCTGGCTGCTGAAGAGTGCAGACTGGCTATGGCGTTACGCCAAG	694
181	LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys	200
695	GACTTCAACCGCTTAAGAGAGAGATGCAGCTCCAGAGCTTCAAGTCAACCTGCACTTG	754
201	AspPheAsnArgLeuLysLysMetGlnProProAlaAlaAlaValThrLeuHisLeu	220
755	GAGGCACATGGTTTC	769
221	GlyAlaHisGlyPhe	225
RESULT 7		
US-09-016-534-2		
Sequence 2, Application US/09016534		
Patent No. 6143874		
GENERAL INFORMATION:		
APPLICANT: CHANG, MING-SHI		
APPLICANT: ELLIOTT, GARY S.		
APPLICANT: SARMIENTO, ULLA		
APPLICANT: SENALDI, GEORGIO		
TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1		
NUMBER OF SEQUENCES: 16		
CORRESPONDENCE ADDRESS:		
ADDRESSEE: AMGEN INC.		
STREET: ONE AMGEN CENTER		
CITY: THOUSAND OAKS		
STATE: CA		
COUNTRY: USA		
ZIP: 91320		
COMPUTER READABLE FORM:		
MEDIUM TYPE: Floppy disk		
COMPUTER: IBM PC compatible		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: PatentIn Release #1.0, Version #1.30		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/09/016,534		
FILING DATE:		
CLASSIFICATION:		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: US 08/792,019		
FILING DATE: 03-FEB-1997		
ATTORNEY/AGENT INFORMATION:		
NAME: COOK, ROBERT R.		
REGISTRATION NUMBER: 31,602		
REFERENCE/DOCKET NUMBER: A-442B		
INFORMATION FOR SEQ ID NO: 2:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 225 amino acids		

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-016-534-2

Alignment Scores:

Pred. No.: 3,22e-105 Length: 225
Score: 1193.00 Matches: 218
Percent Similarity: 98.22% Conservative: 3
Best Local Similarity: 96.89% Mismatches: 4
Query Match: 79.64% Indels: 0
DB: 4 Gaps: 0

US-09-931-704-4 (1-819) x US-09-016-534-2 (1-225)

Qy 95 ATGGACCTCCGAGGAGGAGCTGCTGGGGATGTTAGCTTGCTATGACGGTGTGTGG 154
Db 1 MetAspLeuArgAlaGlyAspSerTrpGlyMetLeuAlaCysLeuCysThrValLeuTrp 20
Qy 155 CACCTCCCTGCAGTCCAGCTCTTAATCGCAGGAGATCCAGCCCTGGCCCTCCATC 214
Db 21 HisLeuProAlaValProAlaLeuAsnA-gThrGlyAspProGlyProSerIle 40
Qy 215 CAGAAACCTATGACCTCACCCCTACCTGGAGCATCAACTCCGCGACTTAGCTGGACC 274
Db 41 GlnLysThrTyAspLeuThrArgTyLeuGluHisGlnLeuArgSerLeuAlaGlyThr 60
Qy 275 TACCTGAATCTAGCTGGGGCCCTTTCAACGAGCCCTGACTTCAATCCTCTCGACTGGG 334
Db 61 TyrLeuAsnTyLeuGlyProProPheAsnGluProAspPheAsnProProArgLeuGly 80
Qy 335 CGAGAACTCTGCCAGGCGCAGGTCAACTTGGAGTGTGGGAGCCCTCAATGACAGG 394
Db 81 AlaGluThrLeuProArgAlaThrValAspLeuGluValTrpArgSerLeuAsnAspLys 100
Qy 395 CTGCGGCTGACCCAGCAATATGAGGGGTACAGTCACCTCTCTGTGTTACTTGGCGCTC 454
Db 101 LeuArgLeuThrGlnAsnTyGluAlaTySerHisLeuLeuCysTyLeuArgGlyLeu 120
Qy 455 AACCTGAGCTGCCACAGCTGAACTCCGAGCTAGCTGCCGCCACTTCTGTACCAGCTC 514
Db 121 AsnArgGlnAlaAlaThrAlaGluLeuArgArgSerLeuAlaHisPheCysThrSerLeu 140
Qy 515 CAGGCGCTGCTGGGAGCATTCAGGTGTCTATGCGAGCGCTTGCTACCCACTGCCCCAG 574
Db 141 GlnGlyLeuLeuGlySerIleAlaGlyValMetAlaLeuGlyTyProLeuProGln 160
Qy 575 CTTCTCCAGGAGTACGAGCAGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCTGGGCGCT 634
Db 161 ProLeuProGlyThrGluProThrTrpThrProGlyProAlaHisSerAspPheLeuGln 180
Qy 635 AAGATGATGACTCTGGCTGCTGAGGAGCTGAGAGCTGAGAGCTGGCTATGGCTTCAGCCA 694
Db 181 LysMetAspAspPheTrpLeuLeuLysGluLeuGlnThrTrpLeuTrpArgSerAlaLys 200
Qy 695 GACTTCAACCGGCTTAAGAAGAGATCGAGCTCCAGAGCTTCAGTCACCTCAGCTTG 754
Db 201 AspPheAsnArgLeuLysLysMetGlnProProAlaAlaValThrLeuHisLeu 220
Qy 755 GAGGCACATGGTTTC 769
Db 221 GlyAlaHisGlyPhe 225

RESULT 8

US-08-642-255-32
Sequence 32, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-32

Alignment Scores:

Pred. No.: 7.32e-06 Length: 330
Score: 146.00 Matches: 74
Percent Similarity: 36.36% Conservative: 2
Best Local Similarity: 35.41% Mismatches: 89
Query Match: 9.75% Indels: 44
DB: 1 Gaps: 11

US-09-931-704-4 (1-819) x US-08-642-255-32 (1-330)

Qy 25 GCGGCTCGCTCCCTCCACTCCGCCAGCC---TCTGGAGAGGAGCGCGCGCGCGCGCC 81
Db 35 AlaProGlyProProGlyProProGlyProProGlyProProGlyAlaProGlyProPro 54
Qy 82 GCGCGCCG---CCCATGGACCTCCGAGCGGCGACTCGTGGGGATCTTAGCTTGCCT 138
Db 55 GlyProGlyProProGlyProPro-----GlyProAlaGlyProValGlySerPro 72
Qy 139 ATGCAGGTGTGTGGCACCTCCCTGCAGTGCCA-----GCTCT 177
Db 73 -----GlyAlaProGlyPro-ProGlyProProGlyProProGlyProProGlyAlaPr 90
Qy 178 TAATCGCAGGAGATCCAGGCGCTGGCCCTTCCATCCAGAAACCTATGACCTACCCG 237
Db 90 oGlyProProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109
Qy 238 CTACTTGAGGATCAACTCCGAGCTTAGCTGGGACCTACTGAACTACTCTGGGGCGCCC 297
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129
Qy 298 TTTCAACGAGCTGACTTCAATCTCTCGACTGGGGCGAGAACTCTGCCAGGGCCAC 357
Db 130 -----GlyProProGlyProProGlyProProGlyProPro 141
Qy 358 GGTCAACTTGAAGTGTGGCGAGGCTCAATGACAGGCTGGCGCTGACCCAGAACTATGA 417
Db 142 GlyProAlaGlyProValGlySerPro----- 150
Qy 418 GCGGTACAGTCACTCTCTGTGTTACTTGGTGGCGCTCAACCGTCAGGCTGCCACAGCTGA 477


```
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-65

Alignment Scores:
Pred. No.: 7.88e-06 Length: 408
Score: 146.00 Matches: 74
Percent Similarity: 36.36% Conservative: 2
Best Local Similarity: 35.41% Mismatches: 89
Query Match: 9.75% Indels: 44
DB: 4 Gaps: 11

US-09-931-704-4 (1-819) x US-08-475-411A-65 (1-408)
QY 25 GCGGCTCGCCTCCACTCCGCGAGCC---TCTGGAGAGGAGCGCGCGCGCGCGCC 81
Db 35 AlaProGlyProProGlyProProGlyProProGlyProGlyAlaProGlyProPro 54
QY 82 GCGCCCGCAG---CCCCATGGACCTCCGAGCAGGAGCTCGTGGGGAGTGTAGCTTGCCT 138
Db 55 GlyProGlyProProGlyProProPro-----GlyProAlaGlyProValGlySerPro 72
QY 139 ATGCACGGTGTGTGGCAGCTCCCTGAGTGCCTCA-----GCTCT 177
Db 73 -----GlyAlaProGlyPro--ProGlyProProGlyProProGlyProGlyAlaPr 90
QY 178 TAATCGCAGAGAGATCGAGCGCTGGCGCTCCATCCAGAAACCTAGACCTCACCG 237
Db 90 oGlyProProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109
QY 238 CTACTCGGAGCATCAACTCCGAGCTTAGCTGGAGCTACCTGAACCTACTCTGGGGCCCC 297
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129
QY 298 TTTCACGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCAC 357
Db 130 -----GlyProProGlyProProGlyProProGlyProProGlyProPro 141
QY 358 GGTCAACTTGAAGTGTGGCGAGCCTCAATGACAGCGTGGCGCTGCGGCTACCCAGAACTATGA 417
Db 142 GlyProAlaGlyProValGlySerPro----- 150
QY 418 GCGGTACAGTCACTCTCTGTTACTTGTGGTGGCTCAACCGTCAGCTGCCACAGCTGA 477
Db 151 GlyAlaProGlyProProGlyProProGlyProProGlyProProGlyProProGly----- 166
QY 478 ACTCCGACGTAGCTGGCCCACTTCTGTACAGCCTCCAGGGCTCGTGGCGAGCATTCG 537
Db 167 AlaProGlyProProGlyPro-----ProGlyProProGlyProPro--GlyPro---Al 183
QY 538 AGGTGTATGGGAGCCTTGGTACCACTGCCCGCCAGCCTCTCCAGGGAGTGA-----CC 594
Db 183 aglyProValGlySerProGlyAlaProGlyProProGlyProProGlyProProGlyPro 203
QY 595 AGCTGGGCGCCCTGGCCCT 613
Db 203 OProGlyAlaProGlyPro 209

RESULT 11
```

US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-029A-65

Alignment Scores:
Pred. No.: 7.88e-06 Length: 408
Score: 146.00 Matches: 74
Percent Similarity: 36.36% Conservative: 2
Best Local Similarity: 35.41% Mismatches: 89
Query Match: 9.75% Indels: 44
DB: 4 Gaps: 11

US-09-931-704-4 (1-819) x US-08-478-029A-65 (1-408)
QY 25 GCGGCTCGCCTCCACTCCGCGAGCC---TCTGGAGAGGAGCGCGCGCGCGCC 81
Db 35 AlaProGlyProProGlyProProGlyProProGlyProProGlyAlaProGlyProPro 54
QY 82 GCGCCCGCAG---CCCCATGGACCTCCGAGCAGGAGCTCGTGGGGAGTGTAGCTTGCCT 138
Db 55 GlyProGlyProProGlyProProPro-----GlyProAlaGlyProValGlySerPro 72
QY 139 ATGCACGGTGTGTGGCAGCTCCCTGAGTGCCTCA-----GCTCT 177
Db 73 -----GlyAlaProGlyPro--ProGlyProProGlyProProGlyProGlyAlaPr 90
QY 178 TAATCGCAGAGAGATCGAGCGCTGGCGCTCCATCCAGAAACCTAGACCTCACCG 237
Db 90 oGlyProProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109
QY 238 CTACTCGGAGCATCAACTCCGAGCTTAGCTGGAGCTACCTGAACCTACTCTGGGGCCCC 297
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129
QY 298 TTTCACGAGCCTGACTTCAATCCTCTCGACTGGGGGCGAGAACTCTGCCAGGGCCAC 357
Db 130 -----GlyProProGlyProProGlyProProGlyProProGlyProPro 141
QY 358 GGTCAACTTGAAGTGTGGCGAGCCTCAATGACAGCGTGGCGCTGCGGCTACCCAGAACTATGA 417
Db 142 GlyProAlaGlyProValGlySerPro----- 150
QY 418 GCGGTACAGTCACTCTCTGTTACTTGTGGTGGCTCAACCGTCAGCTGCCACAGCTGA 477
Db 151 GlyAlaProGlyProProGlyProProGlyProProGlyProProGlyProProGly----- 166
QY 478 ACTCCGACGTAGCTGGCCCACTTCTGTACAGCCTCCAGGGCTCGTGGCGAGCATTCG 537
Db 167 AlaProGlyProProGlyPro-----ProGlyProProGlyProPro--GlyPro---Al 183
QY 538 AGGTGTATGGGAGCCTTGGTACCACTGCCCGCCAGCCTCTCCAGGGAGTGA-----CC 594
Db 183 aglyProValGlySerProGlyAlaProGlyProProGlyProProGlyProProGlyPro 203
QY 595 AGCTGGGCGCCCTGGCCCT 613
Db 203 OProGlyAlaProGlyPro 209

Db 73 -----GlyAlaProGlyPro-ProGlyProProGlyProProGlyProProGlyAlaPr 90
QY 178 TAATCGACAGGAGATCCAGGCGCTGCGCCCTCCATCCAGAAAACCTATGACCTCACCCG 237
Db 90 oGlyProGlyProProGlyPro--ProGlyProProGlyProAlaGlyProValGly 109
QY 238 CTACTGGAGCACTACCTCGCAGCTTAGCTGGAGCCTACTGAACCTACTGGGCGCCCG 297
Db 110 SerProGlyAlaProGlyProProGlyProProGlyProProGlyProProGlyAlaPro 129
QY 298 TTTCAACGAGCCTGACTTCAATCCTCTCGACTGGGGGAGAACTCTGCCAGGGCCAC 357
Db 130 -----GlyProGlyProProGlyProProGlyProProGlyProPro 141
QY 358 GGTCACTTTGGAAGTGTGGAGAGCCTCAATGACAGGCTGCGGCTGACCCAGAACTATGA 417
Db 142 GlyProAlaGlyProValGlySerPro----- 150
QY 418 GCGGTACAGTCACTCTGTTACTTGTGCTGGCCTCAACCGTCAGGCTGCCACAGCTGA 477
Db 151 GlyAlaProGlyProProGlyProProGlyProProGlyProProGly----- 166
QY 478 ACTCGACGTAGCTGGCCACCTTCTGTACACGCTCCAGGCGCTGCTGGGAGCAATGC 537
Db 167 AlaProGlyProProGlyPro-----ProGlyProProGlyProPro-GlyPro--Al 183
QY 538 AGGTGTATGGCGACCTTGGTACCCACTGCCCCAGCCTCTGCCAGGGAGCTGAG---CC 594
Db 183 aglyProValGlySerProGlyAlaProGlyProProGlyProProGlyProProGlyPr 203
QY 595 AGCTGGGGCCCTGGCCCT 613
Db 203 oProGlyAlaProGlyPro 209

RESULT 12

US-08-642-255-62
; Sequence 62, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 27299 FHT UR
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1064 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-62
Alignment Scores:
Pred. No.: 3.65e-05 Length: 1064
Score: 140.50 Matches: 81
Percent Similarity: 37.95% Conservative: 4
Best Local Similarity: 36.16% Mismatches: 94
Query Match: 9.38% Indels: 45
DB: 1 Gaps: 15
US-09-931-704-4 (1-819) x US-08-642-255-62 (1-1064)
QY 25 GCGGCTGCGCCCTCCGACCTCCGCGGAGGAGGAGCGCGCGCGCGCGCGCGCGCGCG 84
Db 125 AlaProGlyProAlaGlyProProGlySer---ArgGly---AspProGlyProProGly 142
QY 85 ---CCCAGAGCCCATGGACCTCCG---AGCAGGGGACTCGTGGGGGATGTTAGCTTGGCCT 138
Db 143 AlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaPro 162
QY 139 ATGACAGGTGCTGTGGACCTCCCTGTCAGTCCAGCTCTTAATCGCACAGGAGATCCAGG 198
Db 163 ---GlyProAlaGlyProProGly-----SerArg-GlyAspProGly 175
QY 199 CCCT-----GGCCCTCCATCCAGAAA-----CCTATGACCT 231
Db 175 yProProGlyAlaProGlyProAlaGly-ProProGlySerArgGlyAspProGly-Pro 194
QY 232 CACCCGCTACCTGGAGCATCAACTCCGCGAGCTTAGCTGGGACCTACCTGAACCTACCTGGG 291
Db 195 ProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGly 214
QY 292 GCCCCTTTCAACGAGCCTGACTCAATCTCCTCGACTGGGGGAGAACTCTGCGCCAG 351
Db 215 AlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaPro 234
QY 352 GGC---CACGTCAACTTGAAGTGTGCGAAGCCTCAATGACAGGCTCGCGCTGACCCA 408
Db 235 GlyProAlaGlyProProGlySerArgGlyAspPro-----GlyProPro 249
QY 409 GAACTATGAGCGGTACAGTCACTCTGTTACTTGGCTGGCCTCAACGCTCAGGCTGC 468
Db 250 GlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAla 269
QY 469 CACAGCTGAACCTCCGAGTAGCCT-----GGCCACTT-----CTG 504
Db 270 HisGly-----ProAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAla 287
QY 505 TACCAGGCTCCAGGCGCTGCTGGGAGCATTCAGGTGTCATGGGAGCGCTTGGGTACCC 564
Db 288 HisGlyProAlaGlyProLysGlyAlaHisGlyProAla---GlyProLys-GlyAlaPr 306
QY 565 ACTGCCCCAGCCTCTGCCAGGAGTAG-----CCAGCTGGGCGCCCTGGCCCC 612
Db 306 oGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaProGlyPr 326
QY 613 TGCC 616
Db 326 oAla 327
RESULT 13
US-08-555-669-12
; Sequence 12, Application US/08555669
; Patent No. 5773248
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds


```

; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: US/08/555,669
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8389-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-555-669-12

Alignment Scores:
Pred. No.:      Length:      684
Score:          Matches:      71
Percent Similarity: 39.07%    Conservative: 13
Best Local Similarity: 33.02% Mismatches:     56
Query Match:     Indels:      75
DB:              Gaps:        16

US-09-931-704-4 (1-819) x US-08-555-669-12 (1-684)
QY 34 CCTCCCACTCGGCAGGCC-----TCTGGGAGAGGA 63
Db 33 ProProGlyProGlyArgProGlyLysProGlyGlnAspGlyIleAspGlyGluAla 52
QY 64 GC CGCGCCGCCGCCGCCGCCCGCAG---CCCCATGGACCTCCGAGCAGG---GGACTC 117
Db 53 GlyProProGlyLeuProGlyProGlyProGlyGlyAlaProGlyLysProGlyLys 72
QY 118 GTGGGGATTAGTTGCCTATGCACGGTGCTGTGGCACCTCCCTGCAGTCCCAGCTCT 177
Db 73 ProGlyGluAlaGlyLeuPro-----Gly-LeuProGlyValAspGlyLe 87
QY 178 T---AATCGCACGAGATCCAGGCCCT-----GGCCCTCCATCCAGAAACTCATGA 228
Db 87 uThrGlyArgAspGlyProProGlyProGlyLysGlyAlaProGlyGlySer-----105
QY 229 CCTACC CGCTACCTGGAGATCAACTCCGAGCTTAGCTGGACCTACCTGAATCACT 288
Db 106 -----Le 106
QY 289 GGCGCCCTTTTCACAGGACCTGACTTCAATCTCCTCGACTGGGGCGAGAACTCTGCC 348
Db 106 uGlyProPro-----GlyProProGlyLeuGlyGlyLysGlyLeu-P 120
QY 349 CAGGGCCACGGTCAACTTTGGAAGTGTGGGAGAGCTCAATGACAGGCTCGGCTGACCCA 408
Db 120 roGlyProPro-----GlyGluAlaGlyValSerGlyPro-----Prog 133
QY 409 GAACATATGAGCGGTACAGTCACTCTGTGTTACTTGTGGTGCCCTCAACGGTCAGCTGC 468
Db 133 lyGlyIle-GlyLeuArgGlyPro-----ProGlyProProGly---- 145
QY 469 CACAGCTGAACCTCGCAGTAGCTCGGCCCACTTCTGTACACAGCCTCCAGGGCCTGCTGG 528

```